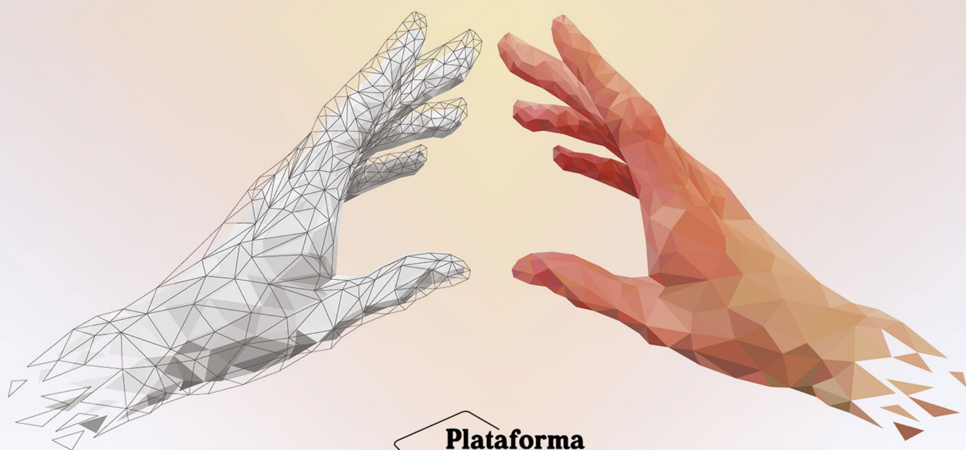
JORDI TORRES

LA INTELIGENCIA ARTIFICIAL EXPLICADA A LOS HUMANOS



Plataforma Actual

«Una lectura imprescindible para cualquiera que desee entender la revolución que representa la IA».

Del prólogo de Mateo Valero, director del Barcelona Supercomputing Center

La inteligencia artificial explicada a los humanos

Jordi Torres

Prólogo de Mateo Valero



Primera edición en esta colección: septiembre de 2023

Plataforma Editorial
c/ Muntaner, 269, entlo. 1ª – 08021 Barcelona
Tel.: (+34) 93 494 79 99 – Fax: (+34) 93 419 23 14
www.plataformaeditorial.com
info@plataformaeditorial.com

ISBN: 978-84-19655-57-8

Diseño de cubierta:
Sara Miguelena

Fotocomposición:
Grafime

A mi familia

Índice

Prólogo

Es un gran honor para mí prologar este libro. En primer lugar, porque está escrito por el profesor Jordi Torres. Jordi es, más que un colega científico, un amigo con quien he podido colaborar durante muchos años en la Universidad Politécnica de Cataluña (UPC), en el Centro Europeo de Paralelismo de Barcelona (CEPBA) y en el Barcelona Supercomputing Center (BSC). Lo conocí a mediados de los ochenta cuando él era aún un estudiante brillante. En esa época se estaba construyendo el departamento de Arquitectura de Computadores de la facultad de Informática de la Universidad Politécnica de Cataluña, del que yo era el director. Solo buscaba primeras figuras y Jordi lo era. Andaba detrás de él para que se quedara como profesor con nosotros. Y según cuenta Jordi, se decantó por quedarse porque se enamoró del aroma de entusiasmo por la docencia y la investigación, libertad y «buen rollo» entre los profesores y los alumnos que transmitía nuestro departamento. ¡Qué suerte tuvimos de que se quedara con nosotros! Ha sido un lujo que nunca le podremos agradecer lo suficiente. Lo cierto es que han sido y son años vibrantes, con ansias de aprender, trabajando muchas horas juntos, con el convencimiento de que estábamos haciendo algo importante para la sociedad como profesores de la UPC y con la creación del CEPBA y luego del BSC. Jordi ha estado siempre allí, como uno más del equipo, jovial, constructivo, atento a los detalles, constantemente atraído por los temas nuevos y apasionado a la hora de explicar nuestros avances a todo el mundo. Es una persona, un amigo, al que admiro por su energía y entusiasmo, por su forma de ser, de hacer y

de ayudar a construir. Muchas veces pienso que ha valido la pena dedicar hasta ahora 49 años como profesor en la UPC por haber tenido la suerte de encontrar y trabajar con personas como él.

En segundo lugar, estoy feliz por saber que el libro va a ser editado por Jordi Nadal, creador de Plataforma Editorial y también mi amigo. De toda la gente que conozco, Jordi es quien más hace por el fomento de la lectura. También es autor de varios libros, pero yo les recomiendo en particular *Libroterapia. Leer es vida*. Y esta idea central de animar y ayudar a la gente a vivir es la que mueve a Jordi a realizar una labor social que nunca le sabremos ni podremos agradecer. Hace realidad, día a día, aquel proverbio hindú que reza: «Un libro abierto es como un cerebro que habla; cerrado, un amigo que espera; destruido, un corazón que llora». O el otro árabe que dice: «Un libro es como un jardín que se lleva en el bolsillo».

Y finalmente, me siento honrado porque el libro habla de inteligencia artificial y de que el gran avance en este campo ha sido posible gracias a la existencia de ideas brillantes de los investigadores, que han podido ser llevadas a la práctica por la existencia de grandes cantidades de datos y de computadores potentísimos que ejecutan estos modelos de redes neuronales que contienen ya billones de parámetros. Nuestro equipo lleva dedicados muchos años al desarrollo de procesadores y supercomputadores potentísimos, así como a técnicas para programarlos de forma que puedan ser usados, como se describe en el libro, para sorprender al mundo.

Tal vez no somos conscientes de que alrededor del año 1950 ocurrieron tres aportaciones a la ciencia que han cambiado nuestro día a día: la invención del transistor, el descubrimiento de la estructura en doble hélice del ADN y la propuesta de la inteligencia artificial.

El transistor, inventado en el año 1947, es la tecnología que, hasta ahora, más ha cambiado la sociedad en menos tiempo. Porque de los transistores salen los procesadores y estos son los responsables del mundo digital que nos envuelve y que ha cambiado todos los aspectos

de nuestra vida. Tuve la suerte de estudiar transistores en la carrera de ingeniería de Telecomunicación en Madrid, que acabé en 1974, y esto me permitió, desde mediados de los setenta, empezar a investigar técnicas para hacer que los procesadores aprovecharan la reducción continua en el tamaño de los transistores (la ley de Moore), y fueran cada vez más rápidos. Y, a partir de aquí, construir sistemas con varios procesadores que colaboraran de forma paralela para mejorar aún más la velocidad de procesamiento. Los supercomputadores son los computadores más rápidos del mundo y se construyen mediante un gran número de procesadores muy rápidos (en la actualidad hasta varios millones) con sus grandes memorias asociadas, conectados a través de una red de interconexión muy veloz.

Los computadores paralelos fueron el embrión de un centro que creamos en el año 1985 en la UPC, que se llamó Centro Europeo de Paralelismo de Barcelona (CEPBA), y que veinte años después, en 2005, permitió la gestación del Barcelona Supercomputing Center (BSC), actualmente un instrumento de primer nivel para la ciencia y la ingeniería al servicio de la sociedad. Fuimos creados para ser hasta 60 personas y en mayo de 2023 participamos más de 850 personas. El BSC es un punto de encuentro entre sus patronos, que son el Gobierno de España, el Gobierno de Cataluña y la UPC, y entre la ciencia y la ingeniería (las ideas) y la sociedad. El BSC es un proyecto colectivo que no sería posible sin su equipo humano, personas muy valiosas que trabajan muchas horas y con mucha ilusión.

Sin duda, un éxito que a menudo describo como el mayor *spin-off* que ha creado una universidad española y que alberga, además del excelente equipo humano, un supercomputador en la capilla de la Torre Girona de la UPC, el MareNostrum. Hoy en día, para hacer investigación de excelencia se necesitan infraestructuras de investigación avanzadas, como un telescopio o un acelerador de partículas. Pero, ante todo, se necesitan supercomputadores, un instrumento de computación de altas prestaciones que ha devenido clave para investigar rigurosamente en

casi todas las ramas de la ciencia. Y entre ellas, como describe de manera magistral Jordi en este libro, la supercomputación también es el gran vector impulsor de la inteligencia artificial (IA), una tecnología que ha avanzado a pasos agigantados en el último decenio, revolucionando e impactando de manera positiva en muchos aspectos de nuestra vida cotidiana, y más que lo hará.

En 1953 se descubrió la estructura del ADN, pero la tecnología no fue capaz de secuenciar genomas hasta pasado el año 2000. A partir de ahí, los avances tecnológicos en el diseño de secuenciadores han permitido no solo reducir su tamaño y precio, sino también aumentar su velocidad, de forma que hoy es posible secuenciar las células en unas pocas horas. Este hecho abre unas posibilidades enormes, como descubrir la influencia de la estructura de nuestros genes en aspectos tales como su relación con el cáncer.

En el año 1956 se reunieron en una escuela de verano unos cuantos investigadores de primer orden que empezaron a discutir sobre las posibilidades de que los computadores, cuyas velocidades aumentaban día a día, pudieran ejecutar programas que simularan algunos aspectos del funcionamiento del cerebro. Allí nació el concepto de «inteligencia artificial». Como se explica en las siguientes páginas, estas técnicas se fueron desarrollando durante muchos años, pero no pudieron demostrar sus enormes capacidades hasta que las acompañamos de la existencia de grandes cantidades de datos y computadores de muy altas prestaciones.

Muchas veces se utilizan los supercomputadores, la información genómica y la inteligencia artificial para avanzar de manera sorprendente en la medicina de precisión (la medicina personalizada). Me gusta explicar cómo se ha logrado predecir el plegamiento de una proteína a partir de la secuencia de aminoácidos, y que para mí es sin duda un avance merecedor del Premio Nobel. Cada vez estoy más convencido de que estamos cerca de que haya Premios Nobel de Medicina o de Literatura que se otorgarán a informáticos. En el BSC trabajamos en investigaciones donde la supercomputación, junto con la

IA, sirven para mejorar la salud de las personas, por ejemplo, para prevenir y curar el cáncer de manera personalizada, o para estudiar y mitigar el cambio climático. Para ello, la IA necesita usar muchos datos, lo cual plantea importantes desafíos éticos que debemos abordar de forma urgente como sociedad. Pero no debemos tener miedo a la IA, sino estar muy atentos y disponer de todas las precauciones que sean necesarias. Y para ello, la forma de lograrlo es empoderar a la sociedad, lo cual requiere una conciencia social para controlar el mal uso de esta tecnología.

El libro que ha escrito Jordi Torres es de gran valor en este sentido, ya que permite que el lector sea consciente de la vertiginosa velocidad a la que avanza la IA y, en consecuencia, comprenda que debemos actuar de manera rápida y descubrir qué podemos hacer cada uno de nosotros para abordar este tema.

Esta es una obra accesible y comprensible, incluso para aquellos que no tienen experiencia previa en el tema, pues el autor evita el empleo de la jerga técnica y se centra en explicar los conceptos más representativos de una manera clara y rigurosa. El libro se estructura en ocho preguntas clave que ayudarán al lector a reflexionar y formarse su propia opinión sobre esta revolución de la IA en la que nos encontramos inmersos, desde si la IA está desplazando al ser humano, hasta si podemos prescindir de ella, pasando por cómo una IA aprende de los humanos, consigue aprender por sí misma o puede llegar a ser creativa. El texto también trata sobre el problema de la falta de sentido común de la IA actual, aborda la ética en el uso de la IA, el impacto social que todo esto puede conllevar, la necesidad de una regulación global y analiza la situación de la soberanía europea en este ámbito.

Yo no concibo un centro de investigación que no conecte con la sociedad. Creo que es una responsabilidad de los científicos explicar a la ciudadanía lo que hacemos en centros como el nuestro y cómo es la tecnología que la rige. Por eso quiero mostrar mi más profundo agradecimiento a Jordi por el esfuerzo que ha realizado para escribir

este libro y plasmar en él un conocimiento de gran valor acumulado durante muchos años.

Estoy convencido de que el lector tiene en sus manos un libro muy útil y de lectura imprescindible para cualquiera que desee entender esta revolución que representa la IA, y estoy seguro de que será una valiosa contribución a la discusión sobre cómo podemos aprovechar esta tecnología para construir un futuro mejor.

Profesor MATEO VALERO
Director del Barcelona Supercomputing Center

Prefacio

A finales de 2022 hubo un punto de inflexión en nuestra relación con la inteligencia artificial (IA) debido, en gran medida, a la aparición de diferentes programas informáticos al alcance de todos los usuarios. Estas IA permiten —a cualquier persona con acceso a Internet— generar textos e imágenes que en muchos casos es muy difícil saber si han sido creados por una IA o por un humano.

Esto avivó un interesante debate público: hacia dónde se dirige la IA y qué consecuencias puede acabar teniendo para la humanidad. Los medios de comunicación se han hecho eco de cómo la IA forma parte de prácticamente todos los aspectos de nuestra vida y de que cambiará el mundo de forma irreversible. Pese a que como sociedad estamos asimilándolo, aún no hay consenso sobre dónde nos llevará la revolución de la IA en la que nos encontramos inmersos.

Las opiniones están muy polarizadas —como sucede últimamente en casi todo—. Por un lado, hay quien cree que la IA es una aliada que podrá aportar soluciones a los grandes retos que se le presentan a nuestra sociedad. Por otro lado, están los que piensan que la IA es una enemiga de la humanidad, quizás por la influencia que ha tenido la ciencia ficción y las distopías con sus máquinas con superinteligencia, generalmente antropomórficas, capaces de superar y rebelarse contra el ser humano.

En cualquier caso, toda herramienta poderosa puede ser beneficiosa o perjudicial, dependiendo de quién la utilice y con qué fines. Es decir, aunque la inteligencia artificial tiene un gran potencial para mejorar

nuestra vida, su uso imprudente puede ser dañino y tener un impacto negativo en la humanidad.

En líneas generales, reinan la inquietud y la confusión entre la población. Veámoslo con un ejemplo sencillo: muchas personas, cuando navegan por Internet o abren aplicaciones en su móvil cada día, no son conscientes de que están usando una IA y que esta condiciona sus acciones. Este desconocimiento los deja totalmente a su merced.

Otro ejemplo es la aparición del chatbot, un servicio de IA gratuito con una capacidad de escritura tan sofisticada que el texto que produce es inquietantemente verosímil, tanto que parece escrito por un humano. La llegada del ChatGPT, por mencionar el más conocido, ha puesto en jaque el modo en que hemos enseñado y evaluado durante decenios a nuestros estudiantes.

El desconocimiento genera confusión, temor, rechazo. Una de las principales causas de este desconcierto (entre el público en general) es que se tiende a utilizar un lenguaje demasiado técnico cuando quien lo explica es un experto en la materia. Sin embargo, cuando quien lo explica es un divulgador, se enfrenta al desafío de transmitir la esencia y perspectiva del tema en un lenguaje claro, lo cual puede resultar difícil, si no imposible. Y esto, sin duda, genera desasosiego y, demasiado a menudo, claudicación.

La IA ha revolucionado —y lo hará mucho más— la forma en que interactuamos con el mundo, por lo que su comprensión resulta necesaria para entender cómo funcionan las relaciones y la sociedad. Con el objetivo de explicar de manera rigurosa pero accesible el funcionamiento y el impacto de la IA, he decidido escribir un libro dirigido al público no especializado.

Como veremos, a pesar de que el progreso de la IA se debe a tres vectores clave (los avances algorítmicos, la disponibilidad de grandes cantidades de datos y la capacidad de computación), ha sido este último el que ha marcado el ritmo de crecimiento de la IA. Por ello, investigar en supercomputación resulta también de gran ayuda para poder explicar

cómo ha evolucionado y cómo puede continuar evolucionando la IA.

La presente propuesta espera ser, precisamente, una explicación rigurosa y accesible para un lector sin conocimientos técnicos y científicos previos. Para ello recurro a generalizar los conceptos —sin faltar al rigor científico— con el objetivo de que el lector no experto en IA encuentre la lectura amena y comprensible. Se apuesta por un lenguaje llano para la descripción de los conceptos fundamentales a la vez que construye un relato cronológico para facilitar su seguimiento —marcado por las efemérides más útiles para el relato—, y apoyándose en una selección de ejemplos lo más familiares y próximos al lector. El resultado de este esfuerzo de generalización ha llevado a definir cuatro paradigmas en los que se basa la IA, que resultarán de gran ayuda para comprender cómo hemos llegado a donde estamos.

El objetivo de este libro es fomentar la reflexión informada y consciente sobre la evidencia de que nos encontramos ya inmersos, sin vuelta atrás, en un nuevo paradigma coevolutivo en el que humanidad y la IA se han embarcado conjuntamente, gestando una interdependencia y cohabitación que exigen respuestas sin demora, porque la IA no esperará al ser humano.

Por ello, este libro intentará responder a ocho preguntas que buscan sintetizar las principales inquietudes en torno a este tema, ordenadas de manera que compongan un relato de todas las etapas que ha vivido la IA para, finalmente, invitar al lector a reflexionar sobre qué tipo de IA queremos y adónde nos debe llevar:

1. ¿Está la IA desplazando al ser humano?
2. ¿Cómo se creó la primera IA?
3. ¿Cómo una IA empezó a aprender de los humanos?
4. ¿Cómo una IA consiguió aprender por sí misma?
5. ¿Cómo una IA ha conseguido ser creativa?
6. ¿Podrá una IA llegar a pensar?
7. ¿Nos debe preocupar el impacto de la IA en su forma actual?

8. ¿Podemos prescindir de la IA?

Creo sinceramente que todos tenemos el derecho a entender la revolución que supone la IA, pues es una de las mayores que la humanidad ha experimentado, y formarnos nuestra propia opinión. De esta manera, contaremos con las herramientas adecuadas a la hora de tomar decisiones sobre los potenciales riesgos y desafíos asociados a la IA, así como también sobre sus posibilidades y ventajas. Y podremos, además, asegurarnos de que la IA evolucione de una manera responsable y sostenible, con un impacto positivo en el futuro de la humanidad.

1.
¿Está la IA desplazando al ser humano?

Nos hemos acostumbrado a hablar con el asistente personal de voz integrado en nuestro teléfono móvil, a personas que llevan implantes electrónicos y a que algunos vehículos circulen sin conductor en determinados entornos controlados. Cada vez más actividades que antes eran exclusivas de los humanos han pasado a estar a medio camino de la esfera de la máquina digital y la esfera humana, una tendencia que parece imparable gracias a los avances en la tecnología que se conoce como inteligencia artificial (IA).

Todo lo que nos rodea se va impregnando de IA

Tomemos como ejemplo el bot conversacional ChatGPT, que la organización OpenAI puso a disposición del público de forma gratuita y a modo de pruebas a finales de 2022, y que ha demostrado ser capaz de procesar y generar respuestas en forma de texto en el mismo idioma (y adaptado al tono y estilo) de las consultas que recibe.

Aunque ChatGPT tuvo un impacto notable e instantáneo, lo cierto es que la IA ya estaba presente en nuestro día a día de manera transversal: en las aplicaciones que usamos en nuestro trabajo, en nuestro ocio con los sistemas de recomendación de las plataformas para películas o series, en el comercial con los asistentes de Amazon, en nuestro hogar, en el transporte, en los coches a través de los asistentes de navegación...

Pensemos en cómo ha evolucionado el Mobile World Congress (MWC) que se celebra en Barcelona desde el año 2006, cuando los teléfonos móviles no tenían pantallas táctiles y se usaban básicamente para llamar o enviar los mensajes de texto —los SMS—. Los *smartphones* no existían y el mercado estaba dominado por marcas como Nokia, Sony Ericsson, Samsung o Motorola. En cambio, en la última edición, la IA se ha convertido en un elemento inseparable de los terminales, que ya tienen más de IA que de teléfonos, pues lo verdaderamente interesante en ellos son los sensores de que disponen para conocernos mejor.

Analicemos, por ejemplo, cómo el teléfono móvil ha desplazado a las cámaras compactas digitales gracias a lo que denominamos fotografía computacional. Esta fotografía, mediante potentes algoritmos que procesan las imágenes digitales, permiten paliar las limitaciones de sus diminutas cámaras y consiguen que las fotografías sean cada vez más espectaculares, liberándonos incluso de la necesidad de utilizar un trípode o luces accesorias.

Ahora bien, esto ha tenido consecuencias colaterales no deseadas. Por ejemplo, es manifiesto que la aparición de la fotografía computacional ha afectado al mercado laboral con el desplazamiento de los profesionales de la fotografía. Hasta no hace muchos años, cuando en el centro de investigación realizábamos eventos, venía un fotógrafo profesional a hacer un reportaje para poder después acompañar la difusión de los resultados del evento. Ahora ocurre en contadas ocasiones, se suple la figura del fotógrafo profesional con uno de los compañeros o compañeras del departamento de medios que toma unas fotografías con un teléfono móvil, las retoca un poco con alguna aplicación de IA del propio dispositivo móvil y en unos segundos las imágenes del evento ya están colgadas en las redes sociales.

Pero las capacidades de procesamiento de imágenes de la IA van mucho más allá que el simple retoque fotográfico. Actualmente ya es posible identificar a individuos específicos en una multitud. Estas habilidades que pueden adquirir las máquinas a través de la IA pueden

estar directamente en conflicto con cuestiones morales y éticas esenciales. ¿Qué pasaría si un sistema de reconocimiento facial diseñado para identificar a unos terroristas se aprovechara para monitorizar a activistas pacíficos o reprimir a minorías étnicas?

Qué entendemos por IA

Parece que hoy en día todo se basa en un enjambre de ordenadores y dispositivos que ejecutan aplicaciones de cualquier tipo, pero ¿cuáles de ellos, o qué parte de ellos, usan o son verdaderamente «inteligencia artificial»?

Todavía no contamos con una definición aceptada de lo que es la IA, ya que es una ciencia nueva, cambiante y experimental. Antes de avanzar, pues, quisiera ofrecer una definición de base de lo que entendemos por IA. De manera superficial, podríamos aceptar definirla como el esfuerzo por automatizar tareas «intelectuales» mediante una máquina, que, gobernada por un algoritmo, es capaz de ejecutar por sí misma funciones que generalmente requieren de la participación de la inteligencia humana.

Ahora bien, al igual que la inteligencia humana es compleja de estudiar y definir —poliédrica y multidisciplinar—, también lo es la IA, pues ha bebido de diferentes disciplinas y áreas del conocimiento tales como las ciencias de la computación, la lógica, las matemáticas, la psicología, la filosofía, la neurociencia, la lingüística y la física.

Desde los años cincuenta del siglo pasado hasta hoy los computadores han sido cada vez más capaces de «hacer cosas», y así el concepto de máquinas «inteligentes» ha ido evolucionando. Una definición que se ha ido manteniendo desde que la propuso John McCarthy en el año 1956 es la de IA como la ciencia y la ingeniería para crear máquinas que se comporten de una forma que llamaríamos «inteligente» si un humano tuviese ese comportamiento. En cualquier caso, para el propósito de este

libro, podemos quedarnos con que «la IA utiliza muchas técnicas diferentes para resolver una gran cantidad de tareas», como señala una de las autoridades mundiales de este campo, la investigadora inglesa Margaret A. Boden.

El progreso de la IA se ve facilitado por otras disciplinas y tecnologías habilitadoras, como la computación en la nube (*Cloud Computing*), Internet de las cosas (IoT, en sus siglas inglesas), conectividad de alta velocidad (5G), robótica, realidad inmersiva (metaverso) o teléfonos móviles. Para este libro se ha considerado no entrar en ellas, dado el alcance y extensión que se requerirían. Hablaremos de la IA entendida como *software*, es decir, en su vertiente de «algoritmo» sin considerar los artilugios y las interacciones con las tecnologías mencionadas anteriormente.

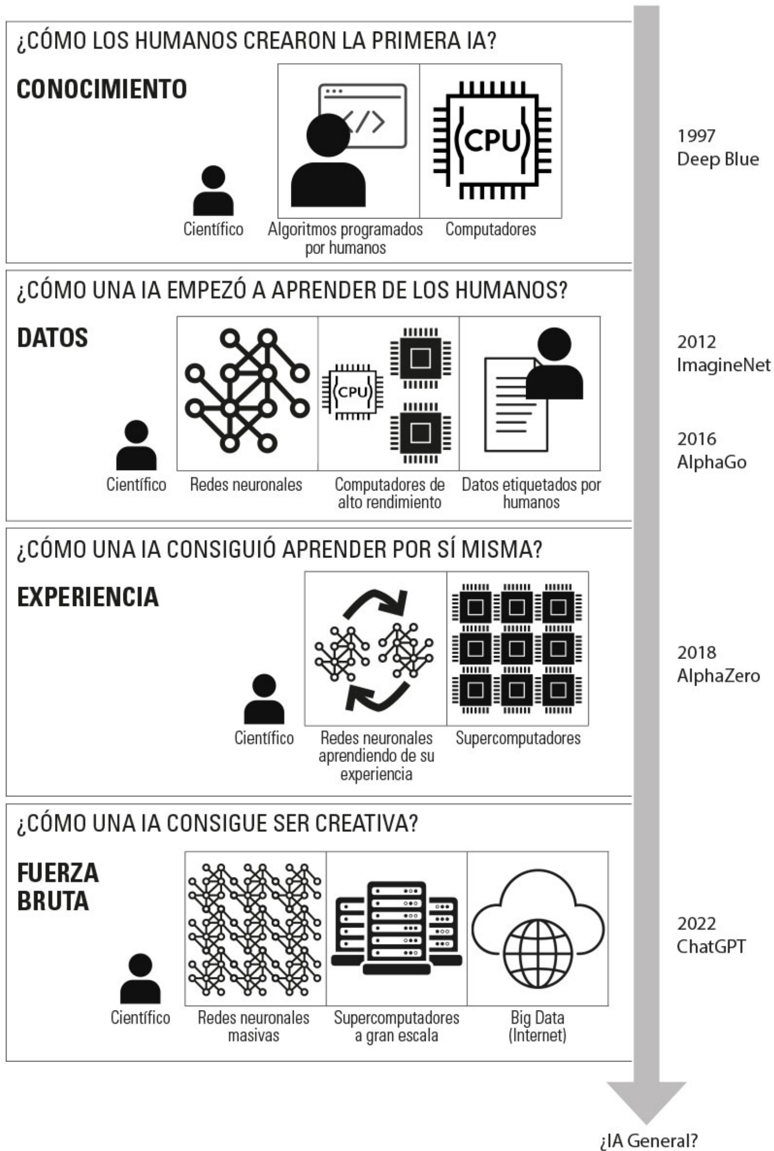
Lo que es seguro es que la IA ya ha penetrado, de una u otra manera, en muchos ámbitos de nuestras vidas para agilizar y optimizar procesos productivos, tanto en nuestro día a día como en lo laboral; en el sector sanitario, la IA es una gran aliada para mejorar los sistemas de diagnóstico o diseño de nuevos fármacos. No cabe duda de que la IA ha venido para quedarse y está viviendo un auténtico *boom*, impregnando todos los aspectos de nuestra sociedad, y continuará haciéndolo, aunque no tengamos claro cómo.

Paradigmas de la IA

Durante el trabajo de generalización y simplificación de conceptos para este libro he llegado a la definición de cuatro paradigmas en que se basa la IA, esquematizados en la figura adjunta: conocimiento, datos, experiencia y fuerza bruta. Han sido configurados por los tres vectores impulsores de la IA (algoritmos, computación y datos), pero definidos de tal manera que realzan el pilar fundamental que ha marcado el ritmo de crecimiento y cambios de la IA: la supercomputación.

La figura puede resultar un tanto críptica a primera vista, pero, a medida que avance la lectura del libro, el lector puede ir revisitándola para observar cómo condensa los conceptos que se van introduciendo y facilita la comprensión del marco general.

El pictograma etiquetado como «científico» que aparece en cada paradigma tiene su intención: recordar al lector que cualquier IA actual no es un ente en sí mismo que pueda pensar y razonar por su cuenta para decidir lo que quiere hacer. Siempre hay una mano humana que está diseñando su propósito y definiendo cómo se vertebra su creación y posterior uso. No sabemos si con los avances podremos llegar a darle una vuelta de tuerca más a la IA hasta que evolucione hacia un ente superior que pueda pensar y razonar por su cuenta.



La IA más allá de la inteligencia humana

A principios de 2023, en plena efervescencia del ChatGPT, en uno de los

programas de radio que dirigimos desde el BSC sobre temas de IA,[1] Xavi Martínez (director del magacín en el que se enmarca el programa) preguntaba si la IA protagonista de «Ahora mismo vuelvo», uno de los capítulos de la serie de *Black Mirror*, era posible. En él, la protagonista decide volcar en una IA todo el contenido que su marido —que acaba de morir— había generado en sus redes sociales, dando vida así a un gemelo digital de su pareja. De esta manera, lo resucita digitalmente y conversa con él como si no hubiera fallecido.

Maite Melero (investigadora experta en modelos de lenguaje) respondió que ya era técnicamente posible con modelos de lenguaje equivalentes a los que dan vida a los bots conversacionales actuales. Pero ¿realmente es una IA que piensa? Melero nos avanza que puede parecérnoslo, dado que «nos resulta alucinante porque el lenguaje nos parece una cosa intrínsicamente humana y estas IA procesan datos y los devuelven como lenguaje».

Los recientes logros de la IA suelen interpretarse como una prueba de que nos encontramos cerca de la llegada de máquinas guiadas por algoritmos que puedan funcionar como una mente humana, llegando a superar a nuestro cerebro. Algunos la llaman superinteligencia. La ciencia ficción es, y ha sido, un buen altavoz de esta visión. Seguramente recuerden la secuencia de la película *Terminator* en la que Skynet toma el control y se vuelve consciente de sí mismo y quiere aniquilar a los humanos porque los considera sus competidores. O la película *2001: Una odisea en el espacio*, en la que la IA HAL 9000, que gobernaba la nave *Discovery*, establece una lucha a muerte con los astronautas humanos para preservar la misión, que era su primer objetivo.

Este es un miedo que está presente desde los inicios de la IA, y más allá de la ciencia ficción. Uno de los padres de la IA y asesor de la película de Kubrick, Marvin Lee Minsky, ya en los setenta decía que «una vez que las computadoras tengan el control, es posible que nunca lo recuperemos. Si tenemos suerte, podrían decidir tenernos como

mascotas».[2] Personajes como Stephen Hawking, Elon Musk, o Stuart Russell, de quien hablaremos más adelante, han expresado sus preocupaciones en esta materia.

Ray Kurzweil (director de Ingeniería de Google) asegura que la aparición de una IA superior a la humana es cuestión de unos pocos decenios, y que esto nos llevará a lo que se conoce por «singularidad tecnológica», momento en el que robots o máquinas dotadas con IA harán absolutamente todo mucho mejor que nosotros.

La hipótesis principal en la que se basan estas predicciones remarca que existe un progreso acelerado en el campo de la IA. Esto plantea una serie de preguntas: ¿tendrá la IA actual un impacto que cambiará la economía, la sociedad y en última instancia la humanidad? ¿Estamos ante un escenario real o de sensacionalismo? ¿Es cierto que la IA está sacando a los humanos de la ecuación? ¿Nos interesa disponer de una inteligencia —artificial— que amplifique la inteligencia humana? ¿O debe preocuparnos?

Ya estamos embarcados en este nuevo paradigma coevolutivo en el que se ha gestado una interdependencia entre la IA y la inteligencia humana. No sabemos hacia dónde nos dirigimos ni cómo terminará el equilibrio humano-IA, aunque la ambición de la IA, como veremos, es colocarse al mismo nivel de la inteligencia humana.

Espero que en este libro el lector pueda encontrar sus propias respuestas a estas preguntas tan importantes para la humanidad. Aunque son respuestas que podemos —y deberíamos— meditar conjuntamente.

En resumen, ¿está la IA desplazando al ser humano?

- Cada vez más actividades que antes eran habilidades exclusivas de los humanos ahora pasan a estar a medio camino entre la esfera de la máquina digital y la humanidad. Estamos ante un nuevo paradigma coevolutivo en el que nos hemos embarcado la humanidad junto con la IA, de modo que se ha gestado una interdependencia cuyo punto de equilibrio no está todavía del todo claro.

- Por IA entendemos el esfuerzo por automatizar las tareas intelectuales, que suelen realizar los humanos, con una máquina que, gobernada por un algoritmo, es capaz de realizar ella sola tareas que generalmente requieren inteligencia humana.
- Los tres pilares fundamentales que impulsan el progreso de la IA son los avances algorítmicos, los datos y la computación.
- Con este libro espero ayudar al lector a que encuentre su propia respuesta a las preguntas planteadas.

2.
¿Cómo se creó la primera IA?

La IA es el resultado de una suma de numerosas inteligencias humanas que se han aunado a lo largo de la historia. Sin embargo, no fue hasta mediados del siglo pasado que los científicos empezaron a visualizar y definir lo que podría llegar a ser. Después de décadas de investigación y desarrollo, a finales del siglo pasado se logró disponer de una IA que tuvo —entre otros— un logro importante para el área: conseguir que una IA venciera al mejor jugador de ajedrez del mundo en aquel momento. Este hito demostró que la IA ya era una realidad tangible llamada a tener un impacto en el futuro de la humanidad.

Máquinas inteligentes

No es sencillo marcar un punto de partida para explicar la historia de la máquina que se convierte en «inteligente». Puede resultar sorprendente, pero sus orígenes están influenciados por contribuciones que datan ya de civilizaciones antiguas. Por poner un ejemplo, hace 2.300 años, Aristóteles planteaba convertir en reglas la mecánica del pensamiento humano. O más próximo en fechas, el polifacético mallorquín Ramon Llull proponía el desarrollo de una máquina llamada *Ars Magna*, que era capaz de realizar demostraciones lógicas para validar o refutar teorías, y que se diseñó como un artefacto mecánico con palancas y ruedas que se movían mediante unas guías que, a pesar de ser muy rudimentaria, fue

el primer intento de utilizar la lógica matemática para producir conocimiento con una máquina.

Desde entonces, la construcción de artilugios dotados de algún tipo de «inteligencia» ha sido constante. Uno de ellos fue el que presentó en 1837 el matemático inglés Charles Babbage, la «máquina analítica», un artefacto diseñado para realizar cálculos matemáticos. En realidad, nunca llegó a terminar su construcción, pero se la considera el primer diseño de un ordenador con verdaderos ingenios mecánicos, basados en el uso de ruedas dentadas y la lógica decimal, e incluía conceptos que usan los computadores actuales. Sin embargo, fue la matemática y escritora británica Ada Lovelace, con la publicación de varios artículos sobre el trabajo de Babbage, la que fue capaz de comprender y prever que las máquinas podían ser útiles más allá que para realizar cálculos matemáticos. Sus notas publicadas contienen lo que hoy se considera el primer algoritmo, uno de los elementos centrales de la IA.

¿Qué es un algoritmo? Podríamos definirlo como una lista de órdenes que se le da a la máquina para que sepa qué hacer, y que cubre todas las posibles opciones a las que se enfrentará en la tarea. Es decir, el algoritmo no hace que la máquina piense, sino que simplemente ejecute. Para ejemplificarlo, podemos considerar que el algoritmo es una receta de cocina que la máquina sigue para obtener el plato deseado.

Los inicios de la IA

Uno de los precursores de la IA fue el matemático e informático teórico Alan Turing, aunque suele ser más recordado como la persona que venció a Enigma, la máquina que transmitía órdenes codificadas a los submarinos alemanes que operaban en el Atlántico durante la Segunda Guerra Mundial.

Alan Turing se mostró siempre muy interesado en la forma de imitar artificialmente las funciones del cerebro humano. Ya 1936, con solo

veinticuatro años, publicó un artículo que hoy se considera el origen de la informática teórica, en el que defendía qué era computable y qué no lo era: lo computable era todo lo que podía resolverse con un algoritmo, lo demás eran tareas no computables. En 1950 publicó el famoso artículo de investigación «Maquinaria computacional e inteligencia», con el que sentó las bases de la IA y propuso un tipo de prueba de imitación, el llamado «test de Turing», para determinar si una máquina puede ser considerada inteligente o no.

El test es un examen que permite determinar la capacidad de una máquina para tener un comportamiento inteligente indistinguible del de un humano. Cuando esto ocurre, puede decirse que la máquina es inteligente. A grandes rasgos, consiste en hacer a una máquina y a un humano ciertas preguntas y a través de sus respuestas saber quién es quién. Si no se logra distinguirlos, significa que el ordenador ha superado la prueba. En la actualidad, es cada vez más habitual encontrarnos con *trabajos* (textos, imágenes, música, distintos tipos de *creaciones*) realizados por una IA que, a priori, nos resultan indistinguibles de las que podría hacer (o ha hecho en el pasado) el hombre.

No obstante, el test de Turing tiene una relevancia controvertida. A pesar de interesar a los filósofos y al público en general durante mucho tiempo, no ha sido importante para la comunidad de la IA en sí misma. Digamos que, por lo general, el objetivo de la IA ha sido proporcionar herramientas útiles, pero no hacer creer a los usuarios que están interactuando con una persona.

Sin duda, Turing comprendió claramente que la inteligencia de la máquina era una extensión lógica y tal vez inevitable de la computación electrónica. Por desgracia, su temprana muerte en 1954 solo le permitió ver la primera generación de ordenadores que existieron, unas máquinas que funcionaban con válvulas —parecidas a las bombillas de ahora— cuya programación se hacía directamente en los propios circuitos y tenía un uso exclusivamente científico y militar. No fue hasta después de su

muerte que, con la aparición de los transistores, se pudieron construir los primeros ordenadores digitales, los cuales dieron lugar a diferentes investigaciones precursoras de la actual IA. De hecho, el término «inteligencia artificial» se acuñó en 1956, atribuido al matemático John McCarthy. La expresión pretendía unificar los diferentes esfuerzos de investigación en cibernética, teoría de autómatas y procesamiento de información compleja para dar a las máquinas la capacidad de «pensar». Esta primera mención tuvo lugar durante una conferencia que duró dos meses en la Dartmouth College de Hannover (Estados Unidos), a la que fueron invitados los principales protagonistas del nuevo campo emergente como Marvin Minsky o Claude Shannon, entre muchísimos otros investigadores.

El campo de la IA progresó rápidamente en los años posteriores a la conferencia de Dartmouth gracias a los avances de los ordenadores digitales. Los programas de financiación para investigaciones, como los gestionados por la Advanced Research Projects Agency de Estados Unidos, permitieron el desarrollo de algoritmos capaces de resolver cada vez más problemas. Fue en ese punto de la historia cuando comenzaron a aparecer laboratorios de investigación en IA de referencia mundial.

Esa situación de bonanza generó una expectativa excesiva entre muchos de los investigadores y se hicieron promesas poco realistas. A mediados de los setenta, cuando comenzó a ser evidente que estas predicciones estaban provocadas por una burbuja de optimismo y que la construcción de sistemas informáticos que mostraran IA sería mucho más difícil de lo esperado, empezaron a languidecer tanto la financiación gubernamental como la de inversores privados. Fue el primer «invierno de la IA», un período de desilusión que, como veremos, no sería el último. Uno de los factores que sin duda llevaron a este invierno fue la poca potencia de los *hardwares* de la época, ya que no podían ofrecer la capacidad de cálculo que se requería para computar los algoritmos de la IA.

Hubo que esperar hasta la década de los ochenta para que se viviera

otra época de bonanza para la IA. Para entonces, los ordenadores comenzaban a ser más potentes y tanto los gobiernos como la industria en general empezaron a mostrar un interés renovado por la tecnología de IA, aunque esta vez centrados en la creación de productos comerciales. En el centro de la tecnología estaba lo que se conoce por «sistemas expertos», cuyo algoritmo tomaba las «decisiones» según un código escrito previamente por un programador. Esta nueva aproximación a la IA se implementó en campos tan variados como el diagnóstico médico, la planificación financiera o el diseño de circuitos microelectrónicos. Es en este contexto cuando se creó la primera IA capaz de vencer al ajedrez al mejor jugador humano del mundo.

Cuando una IA ganó al humano en el ajedrez

El ajedrez es un juego de estrategia muy popular en el que dos jugadores se desafían ante un tablero cuadriculado de 8 x 8 casillas y dos grupos de figuras, 16 para cada jugador. Tengo un recuerdo muy vivo de mi niñez y adolescencia siguiendo absorto cómo se enfrentaban decenas de jugadores en el torneo de Fiesta Mayor de verano de mi pueblo (Argentona, en la provincia de Barcelona).

Durante muchos siglos, el ajedrez fue visto como una manera de medir y comparar la «inteligencia» de las máquinas con la inteligencia humana. En el siglo XX, el matemático Claude Shannon, pionero en la teoría de la información y los fundamentos de las señales digitales, publicó un artículo —que hoy es considerado pionero sobre la programación de computadoras para jugar al ajedrez— en el que presentó estrategias que aún se utilizan en los programas de ajedrez modernos. La dificultad radica en que la infinidad de movimientos posibles hace que no haya dos partidas iguales. Claude Shannon calculó en su artículo que el número de posibles partidas que se pueden jugar al ajedrez son alrededor de 10 elevado a 120. Para un ser humano,

memorizar todos los movimientos de esta cantidad de partidas es algo prácticamente utópico.

A mediados de 1990 se planteó seriamente la posibilidad de diseñar un algoritmo que pudiera adelantar jugadas y los potenciales movimientos en una partida, algo imposible para un humano, incluso para la mente del ajedrecista más avanzado del mundo. La empresa IBM recogió el guante y desarrolló un sistema al que bautizó Deep Blue, una combinación de *hardware* y *software* que tenía el propósito de ganar al mejor jugador de ajedrez humano de ese momento.

Deep Blue era un potente computador diseñado específicamente para jugar al ajedrez. Para construirlo, IBM tuvo que fabricar nuevos chips capaces de trabajar en paralelo, lo que permitía al algoritmo disponer de la evaluación de alrededor de 200 millones de posiciones de tablero por segundo. El algoritmo del Deep Blue, basado en un sistema experto, analizaba de antemano los posibles movimientos y los evaluaba para centrarse en los que parecían más prometedores (los que tenían más probabilidad de ganar).

En este caso, el sistema experto incorporaba en el algoritmo el conocimiento (el código) escrito por expertos humanos. Para crear el sistema Deep Blue, además de ingenieros programadores, IBM contó con la colaboración de grandes maestros del ajedrez que ayudaron a desarrollar el algoritmo. Uno de ellos fue Miguel Illescas (el mejor ajedrecista de España en ese momento) que, además de ayudar a programar la base de datos de aperturas del algoritmo, jugó contra la primera versión de Deep Blue durante la Conferencia Internacional de Supercomputación (ICS, por sus siglas en inglés) en julio de 1995, organizado en Barcelona por el departamento de Arquitectura de Computadores en la UPC.

Era la primera partida de ajedrez a escala mundial entre el computador Deep Blue y un campeón de ajedrez, por lo que la expectación era intensa. Mateo Valero (actual director del BSC) aún cuenta: «Me jugué 100 pesetas de las de entonces con nuestro amigo

Ron Perrott, profesor de la Universidad de Belfast, a favor de Miguel».

Entonces, ¿quién ganó? Aunque Deep Blue perdió contra Illescas, en realidad ganó mucho con su paso por Barcelona: los ingenieros de IBM se valieron de esta experiencia para mejorarlo. Su siguiente enfrentamiento fue contra el vigente campeón del mundo, Garri Kaspárov, en febrero de 1996. De un total de seis partidas, Deep Blue perdió dos juegos a cuatro. Finalmente, el 11 de mayo de 1997, después de una nueva mejora, derrotó a Kaspárov en un *match* de seis partidas, ganando tres juegos y empatando uno.

Sin duda, la victoria de Deep Blue tuvo una gran repercusión mediática. La partida fue bautizada como «el hombre contra la máquina». La IA había salido de los laboratorios para saltar al debate social, y entró en el imaginario colectivo como algo real y tangible, y no ciencia ficción.

En resumen, la IA de Deep Blue se basaba en la incorporación al algoritmo del conocimiento de grandes maestros de ajedrez. Es decir, los humanos enseñaron a la IA las reglas y técnicas de un juego inventado por ellos mismos.

La IA basada en el conocimiento

El sistema experto Deep Blue usaba técnicas de ciertas ramas de la IA que, concediéndonos la licencia de generalizar términos, se engloban como IA clásica o paradigma de una IA basada en el conocimiento.

Un sistema experto como el presentado utiliza una combinación de técnicas de IA, como la lógica simbólica y la inferencia automática, para crear algoritmos que simulen el razonamiento humano. Esto se conoce como IA simbólica, que es una rama de la IA. En estos sistemas se utiliza un lenguaje simbólico para representar el conocimiento. Por ejemplo, la base de conocimiento de Deep Blue contenía información sobre aperturas de ajedrez, estrategias y tácticas obtenidas de los grandes

ajedrecistas.

A partir de esta información, para inferir nuevos conocimientos, se aplican reglas lógicas, que se codifican en el algoritmo como una serie de sentencias condicionales: es decir, una acción y su respuesta están vinculadas por una condición mediante la expresión «si-entonces» («*if-then*», en inglés). Estas sentencias lógicas, que le indican al algoritmo que «si sucede esto, haz esta acción», son definidas de antemano, una por una, por expertos humanos. Este tipo de regla se utiliza aún hoy en la programación de algoritmos.

Para mejorar el rendimiento de Deep Blue se utilizó también una combinación de heurísticas (regla que se utiliza para tomar decisiones valorando cuán buena es o no una situación) para evaluar la calidad de las posiciones de tablero (una técnica dentro de la rama IA numérica).

Deep Blue no solo se nutría de las experiencias de los jugadores humanos, sino que también se benefició de mantener un histórico de partidas jugadas. Esto se hizo con técnicas de *machine learning* —aprendizaje automático—, que permite desarrollar algoritmos que aprenden de manera automática a partir de los datos que recoge a medida que funciona. Es decir, en este caso el algoritmo no cuenta con un código que le brinde todas las opciones posibles, sino que se codifica (se le *instruye*) para que descubra patrones y relaciones en los datos por sí mismo, de manera automática.

Los sistemas informáticos basados en sistemas expertos se han convertido paulatinamente en componentes habituales del diseño del *software* y en muchos casos ya no se etiquetan como «inteligencia artificial», puesto que se han convertido en herramientas habituales para resolver problemas con una base lógica de manera eficiente. Los sistemas de piloto automático de los aviones a reacción son un ejemplo de ello, pese a que hoy en día estos sistemas cuentan con técnicas más avanzadas y son más «inteligentes».

La victoria de Deep Blue, no obstante, en el fondo mostró lo limitados que son los sistemas que codifican a mano el conocimiento. IBM había

gastado millones de dólares y años de tiempo desarrollando un superordenador *solo* para jugar al ajedrez. Deep Blue era lo que conocemos como una IA Específica, ya que solo era capaz de jugar al ajedrez y ni siquiera podía emprender juegos mucho más sencillos, como el tres en raya.

Durante la primera década del nuevo milenio sufrimos otro invierno de la IA. No fue hasta principios de 2010 cuando explosionó nuevamente el interés —y las inversiones millonarias— por el área de la IA.

En resumen, ¿cómo los humanos crearon la primera IA?

- A mediados del siglo XIX Ada Lovelace escribió el que se considera el primer algoritmo y, por ello, se la recuerda como la primera programadora de la historia.
- Un algoritmo es una lista de órdenes en la que se explica a la máquina qué hacer y cómo hacerlo. Se cubren todas las posibles opciones a las que se pueda enfrentar.
- El juego del ajedrez se vio durante años como una herramienta para medir y comparar la «inteligencia» de las máquinas con la humana.
- Alan Turing, a mitad del siglo XX, sentó las bases de la IA.
- La historia de la IA ha vivido períodos de alta exuberancia y rápido progreso, seguidos por períodos de desilusión y poca inversión llamados «inviernos de la IA».
- Los humanos crearon la primera IA con el paradigma de una IA basada en el conocimiento, que consiste en que el conocimiento de expertos humanos fuera introducido «a mano» en el código del algoritmo.
- En 1997 la empresa IBM desarrolló el sistema experto llamado Deep Blue, una IA que ganó al mejor jugador de ajedrez de ese momento, Garri Kaspárov.
- Desde finales de los noventa y principios de la década del 2000 el mundo entró en otro «invierno de la IA» debido a que los sistemas

basados en el conocimiento eran muy limitados.

3.
¿Cómo una IA empezó a aprender de los humanos?

La IA había alcanzado un punto muerto: los sistemas basados en el conocimiento dependían de la pericia de los programadores para alimentarlos con información, pero desperdiciaban todo el conocimiento latente —en los datos disponibles— que no fuera detectado y programado por los ingenieros. Eran excesivamente dependientes y costosos. Así, el siguiente gran paso de la IA fue en esta dirección con la llegada de un nuevo paradigma: la IA basada en datos, fundamentada en el uso de las redes neuronales artificiales.

La IA basada en datos

El paradigma de una IA basada en el conocimiento funciona muy bien si se sabe cómo comunicarle con precisión las reglas y su secuencia. Sin embargo, no funciona tan bien cuando es difícil especificar el algoritmo necesario para resolver un problema. Por ejemplo, si quisiéramos programar una IA para detectar gatos en imágenes, nos resultaría difícil escribir las directrices que permitiesen a la IA cubrir todas las situaciones potenciales, pues los gatos pueden aparecer de múltiples maneras en una imagen y siempre habría casos no considerados en los que esta IA fallaría.

Es entonces cuando aparece el denominado «paradigma de una IA

basada en datos». En esta ocasión le proporcionamos datos a la IA, así como las herramientas para que pueda aprender de estos datos. Por ejemplo, para el reconocimiento de gatos en una foto, introduciríamos una gran cantidad de imágenes en la IA y le indicaríamos si contienen gatos o no. Luego, dejamos que la IA los estudie de manera autómata, con las herramientas de aprendizaje que le hemos proporcionado, para que sea ella la que aprenda a descubrirlos y diferenciarlos de, por ejemplo, otros animales. Es decir, con el paradigma con el que se trabajaba en los años de la creación de Deep Blue, habríamos instruido a la máquina con la siguiente información: si sobre el sofá hay una figura con esta forma, con bigotes y cola, de tal color, entonces es un gato. Por supuesto, esta descripción debería ser extremadamente precisa y solo serviría para una foto. Si en una segunda imagen el gato hubiera girado la cabeza hacia la izquierda, por más que fuera medio centímetro, la IA necesitaría esa información para encontrarlo. Es similar a lo que ocurría a la hora de prever todos los movimientos posibles de una partida de ajedrez: no es un número infinito, pero sí inimaginable. En el nuevo paradigma, en cambio, alimentamos a la IA con, por ejemplo, un millón de fotos y le damos una verdad absoluta: en estas hay gatos y en las restantes, no. A partir de esta verdad dejamos que descubra por sí sola por qué esto es así (con el objetivo de que luego pueda usar esas conclusiones con una nueva imagen).

En otras palabras, le decimos a la máquina qué hacer en lugar de cómo hacerlo. En este caso, se trata de una IA que es capaz de aprender de los datos que tenemos los humanos (ya clasificados en fotos con gato o sin gato), pero que no requiere que los humanos programemos el conocimiento en su algoritmo interno.

Es una técnica de aprendizaje automático que se basa en el uso de redes neuronales artificiales —a partir de ahora nos referiremos a ella solo como redes neuronales—, y es este el paradigma de IA que prevalece en la mayoría de las áreas de investigación y desarrollo en IA. Como veremos con más detalle, es un tipo de algoritmo que se inspira

en la forma en que un cerebro animal conecta las neuronas e intercambia pulsos.

Los orígenes de este tipo de aprendizaje automático vienen de lejos. Se basa en la escuela de pensamiento y en el trabajo de Frank Rosenblatt sobre redes neuronales, que desarrolló en la década de 1950, y cuyos seguidores creían que un sistema inteligente debía moldearse a partir del cerebro y que podía utilizar componentes inspirados en las neuronas biológicas.

El aprendizaje, entonces, se ha convertido en la capacidad más importante de la IA. Se concluyó que si se pudiera generar un sistema para aprender de manera eficiente de los datos, la IA podría extraer y adquirir conocimientos que sus creadores no tuvieran.

Podemos imaginar este tipo de IA como si fuera un niño pequeño que está aprendiendo a reconocer cosas nuevas de su entorno. Al principio, el niño no sabe nada sobre el objeto que queremos que aprenda a identificar, pero le mostramos muchos ejemplos del objeto y le decimos qué es ese objeto. A medida que el niño ve más ejemplos del mismo objeto, el niño va identificando patrones y características que se repiten en el objeto y le permiten con ello aprender a reconocerlo.

Lo mismo ocurre con una IA basada en redes neuronales. Alimentamos (mostramos) a la IA con muchos ejemplos de datos y le decimos qué son cada uno de ellos. La IA utiliza estos ejemplos para identificar patrones y características comunes de estos datos. A medida que recibe más datos, mejora su capacidad para encontrar estos patrones e identificar de qué se trata.

Pilares de la IA

A principios de la década de 2010 los tres vectores impulsores de la IA (datos, algoritmos y computación) presentaron importantes avances, lo que propició un nuevo progreso de la IA. Una parte importante de este

avance había comenzado a gestarse en décadas anteriores con el fenómeno del «Big Data». A mediados de los años noventa se popularizó el uso de Internet y surgió la primera generación de la web. Amazon (1994), Yahoo (1995) y Google (1998), entre otras, ya daban por entonces servicio a un amplio abanico de usuarios que se conectaban desde cualquier lugar del mundo para buscar información, comprar productos o simplemente navegar, mientras generaban datos sin ser conscientes de ello. A finales de los noventa y principios del nuevo milenio, nació una segunda generación de servicios y aplicaciones web, la Web 2.0, en la que los usuarios ya no eran solo consumidores de información sino también productores de una información que, a su vez, pasaba a ser consumida por otros usuarios. Es el inicio de las redes sociales, los blogs, las wikis y la explosión de los dispositivos móviles —teléfonos inteligentes, ordenadores portátiles, *tablets*, etc.—. El último eslabón, hasta el momento, es lo que se conoce como «Internet de las cosas», objetos con sensores conectados a Internet que recopilan datos para que podamos interactuar con ellos.

Pero volvamos al año 2000, cuando, junto con el imparable crecimiento del torrente de datos que circulaban a través de la red, los desarrolladores avanzaban en el diseño de nuevos chips con nuevas (y más potentes) capacidades de computación. Gracias a la aparición de los chips GPU (Graphics Processing Units), unos chips rápidos que la firma Nvidia había creado para tarjetas gráficas de videojuegos en 3D, se pudo desarrollar un nuevo tipo de red neuronal. GPU es un tipo de *hardware* especializado que se utiliza para aumentar la velocidad de ejecución de aplicaciones que requerían realizar muchos cálculos numéricos (como ocurre con los videojuegos).

A diferencia del procesador central de un computador de los del anterior paradigma, un chip acelerador GPU puede procesar múltiples cálculos matemáticos de manera simultánea, lo que permite agilizar el proceso de entrenamiento de una red neuronal y hacer que el aprendizaje sea más rápido.

Así llegamos al año 2012, cuando el Big Data y los procesadores GPU confluyeron en un momento decisivo en la historia de la IA: la competición anual de visión por computador, la ImageNet (Large Scale Visual Recognition Challenge). El desafío consistía en que los equipos participantes diseñaran un algoritmo que pudiera clasificar correctamente, en mil categorías distintas, más de un millón de imágenes de una inmensa base de datos de fotografías. Uno de esos equipos fue el de la Universidad de Toronto, compuesto por Ilya Sutskever y Alex Krizhevsky y dirigido por Geoffrey Hinton.

Hasta ese momento, en la competición se habían utilizado técnicas de programación informática tradicionales, pero ese año el equipo de la Universidad de Toronto abordó el problema con un algoritmo basado en redes neuronales. Entrenaron su red neuronal, llamada AlexNet, con dos GPU. Ganaron de calle la competición. Sus resultados fueron la prueba inequívoca de que este tipo de IA había evolucionado hacia una tecnología práctica y estaba lista para ser aplicada a muchos ámbitos.

A partir de entonces se empezaron a usar chips GPU para entrenar redes neuronales en todos los grupos de investigación en IA. Esto contribuyó a avances importantes en áreas como la visión por ordenador con interpretación de imágenes y vídeos, el procesamiento del lenguaje natural o la comprensión del habla, tres de las principales áreas de investigación y desarrollo de la IA en la actualidad.

El cambio es significativo. Las redes neuronales permiten crear algoritmos muy flexibles para abordar tareas difíciles. Además, ofrecen soluciones menos centradas en la ingeniería con el objetivo de diseñar métodos de propósito general, cuya calidad no está limitada por la comprensión o visión humana del mundo (que antes estaba incluida en el algoritmo de la IA a través del código escrito por los hombres). Ahora, la responsabilidad de la toma de decisiones y del aprendizaje ya no es de los programadores, sino de la propia IA.

Este avance fomentó la inversión en investigación y la llegada de un nuevo ciclo de bonanza para la industria de la IA. Empresas como

Google (Alphabet), Facebook (Meta), Amazon, Apple, Microsoft, Tesla, Baidu, Tencent y Alibaba deben su éxito a que pusieron las redes neuronales en el centro de sus productos y de su modelo de negocio.

La IA vence al humano en el juego del go

Ya hemos visto, con el ajedrez, que el juego ha sido una forma de medir los avances en IA. Para el nuevo paradigma de IA basada en los datos había que buscar un nuevo territorio de combate, y esta vez se escogió el juego del go.

El go es un antiguo juego de mesa que se originó en China hace más de 2.500 años. Consiste en un tablero de 19 x 19 líneas, con 361 posiciones, en el que dos jugadores colocan fichas de dos colores, blancas y negras, y compiten por controlar la mayor cantidad de territorio posible en el tablero. Cuando una ficha se coloca en el tablero ya no se puede mover, pero es posible capturarla al rodearla con las fichas del oponente. El juego termina cuando no hay más lugares en el tablero para colocar fichas (o hasta que uno de los jugadores se rinda). El jugador con más territorio controlado al final del juego gana. Pese a parecer sencillo, es en realidad un juego muy complejo que requiere mucha estrategia y habilidad. El objetivo final es controlar territorio (y no tanto capturar fichas del oponente, aunque sea una parte importante de la estrategia). La cuestión es que los posibles movimientos son tantos que computarlos es una tarea titánica.

Intentar crear reglas como se hizo en Deep Blue resultaba imposible, dada la diversidad de situaciones que pueden aparecer. Incluso se había asumido que una IA jamás superaría al humano en el juego del go, hasta que en 2016 fue conquistado por una IA llamada AlphaGo, principalmente basada en datos.

AlphaGo fue desarrollada por DeepMind (fundada en 2010 por Demis Hassabis, Shane Legg y Mustafa Suleyman), una empresa de IA con sede

en Londres y propiedad de la matriz de Google Alphabet. El equipo de ingenieros de DeepMind enseñó a la red neuronal de AlphaGo a identificar las jugadas victoriosas de aquellas que no lo eran. Lo hicieron a partir de un gran conjunto de partidas reales jugadas por humanos. Para complementar su aprendizaje se utilizaron técnicas clásicas, y la técnica aprendizaje por refuerzo para buscar las mejores jugadas sin tener que calcular todas las permutaciones posibles en cada jugada, sino solo las más pertinentes.

AlphaGo marcó un nuevo hito para la IA al vencer al campeón mundial de go (Lee Sedol). Y, quizás, también en el ámbito de la geopolítica. Según Kai-Fu Lee, escritor experto en IA, la derrota del campeón mundial en el juego predilecto de China fue la alerta para que el Gobierno chino decidiera impulsar los esfuerzos de su país en convertirse en una superpotencia de la IA. Desde entonces, China ha emprendido una carrera por el liderazgo en la IA en la que está avanzando muy rápido.

Las redes neuronales por dentro

Las redes neuronales se inspiran en la neurona biológica. Esto puede conducir a confusiones sobre los paralelismos entre las redes neuronales utilizadas en la IA y el cerebro humano. El cerebro humano es, sin duda, un sistema complejísimo con aproximadamente cien mil millones de neuronas y una mayor cantidad de sinapsis que permiten la conexión entre neuronas. Para más inri, su complejidad no surge simplemente de la conectividad a gran escala de neuronas, sino de una complicada mezcla de procesos químicos y eléctricos que todavía no se acaban de conocer en detalle.

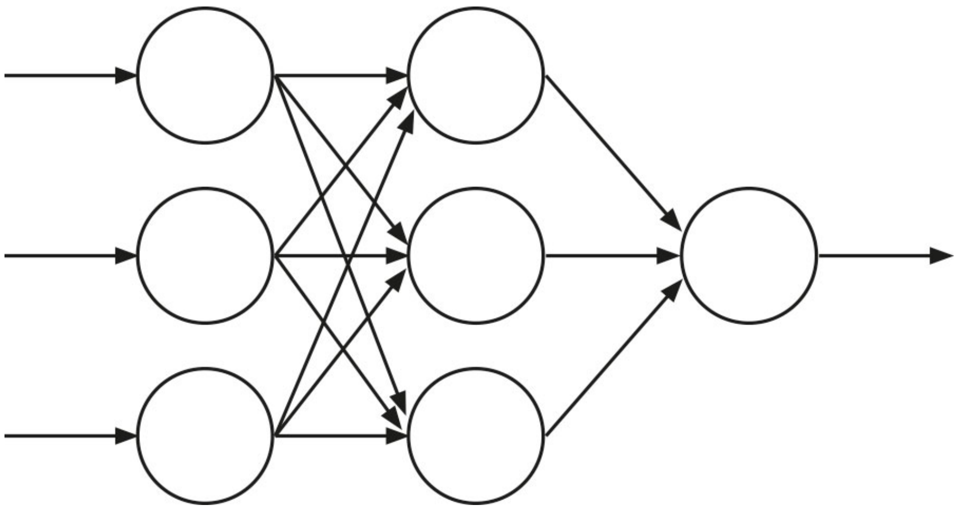
Por lo tanto, creo necesario enfatizar que una red neuronal nada tiene que ver con el cerebro humano, tampoco se intenta que simulen lo que ocurre en el cerebro humano, aunque la neurona biológica y sus

interconexiones sí han servido como inspiración para las redes neuronales artificiales.

Para entender un poco el mecanismo de las neuronas podríamos decir que una neurona biológica consta de tres partes principales: el cuerpo celular, donde reside el núcleo; numerosos filamentos, conocidos como dendritas, que llevan señales eléctricas entrantes en el núcleo; y un filamento único, llamado axón, con el que la neurona transmite una señal de salida al resto de neuronas con las que está conectada. Tanto las dendritas como el axón suelen ramificarse ampliamente; a veces las dendritas reciben estimulación eléctrica de decenas de miles de otras neuronas. Cuando las señales que llegan a través de las dendritas activan la neurona, esta entrega una carga eléctrica saliente a través del axón.

De forma similar, una neurona artificial tiene un núcleo que recibe como entrada la energía de las neuronas con las que está conectada. Si la suma proporcional de la energía que le llega de otras neuronas supera cierto umbral, es decir, si la neurona se activa, envía energía a través de las conexiones de salida que tiene con otras neuronas. Podríamos encontrar una cierta similitud con las sinapsis cerebrales, no obstante, en una neurona artificial, la fuerza se expresa como un número, por lo que cuanto mayor sea el peso de una conexión, mayor será la probabilidad de que la neurona que recibe la señal de energía se active.

Imaginemos que existen millones de estas neuronas artificiales con conexiones entre ellas. Cada una de las conexiones de entrada dispone de una válvula (llave de paso) que permite controlar el flujo de energía que llega a la neurona. Ajustando estas válvulas, es posible regular directamente la influencia que tienen otras neuronas conectadas sobre una neurona en particular. Una vez ajustadas estas válvulas, podemos considerar que la red neuronal ha adquirido el conocimiento y puede realizar la tarea que se le ha encomendado. Un esquema sencillo lo representaría de este modo:



Para referirnos a la organización de las neuronas de una red neuronal usamos el concepto de «arquitectura de una red neuronal». Estas redes neuronales están conformadas por un número concreto de neuronas organizadas y conectadas entre ellas, como podemos observar en la figura: siete neuronas organizadas en tres capas.

Al empezar a diseñar una IA basada en redes neuronales, el primer paso es elegir la arquitectura más adecuada para el propósito encomendado. Crear nuevos tipos de arquitecturas de redes neuronales es uno de los campos de investigación más activos en la IA, puesto que la arquitectura de la red es determinante para conseguir que sea efectiva para el propósito que se persigue.

Una vez hecho esto, se pasa al proceso de «entrenamiento» para que aprenda a realizar alguna tarea concreta. Es esencialmente una cuestión de ajustar los parámetros (también llamados «pesos») de las conexiones entre neuronas, mostrando a la red neuronal datos de los que sepamos la respuesta correcta. Es decir, se examina cuál es el valor que calcula la red neuronal por cada dato de entrada y este valor se compara con la respuesta que se sabe que es el valor correcto y, de esta forma, el sistema de entrenamiento va ajustando los parámetros de las conexiones con un complejo mecanismo matemático.

Volvamos al ejemplo de las fotografías con gatos para tratar de ilustrar cómo podemos entrenar una red neuronal. Supongamos que le damos a la red neuronal muchas imágenes de gatos y sin gatos y le pedimos que intente determinar en cuáles imágenes hay un gato y en cuáles no. A medida que la red neuronal va viendo imágenes y comete errores, vamos haciendo algunos ajustes en los parámetros —usando un mecanismo de optimización matemático que permite que todos los parámetros de la red se ajusten colectivamente, de forma que la próxima vez tenga más probabilidades de adivinar correctamente—. Este proceso de afinar los parámetros de las conexiones entre las neuronas para que todas ellas converjan poco a poco en los valores adecuados se hace de forma iterativa hasta que la red neuronal es capaz de diferenciar entre imágenes con gatos y sin gatos. Para llegar a este punto es necesario ajustar colectivamente muchos parámetros; este es el motivo por el que en la fase de entrenamiento se realizan grandes cantidades de cálculos y se requiere mucha capacidad de computación.

Antes de continuar, pongamos atención a un detalle relevante: en el proceso de entrenamiento disponemos de la respuesta correcta que esperamos que la red encuentre. Esto implica que todos los datos de entrada se han tenido que «etiquetar» con la respuesta correcta. Es una tarea muy tediosa que realizan operadores humanos.

Una vez finalizada la fase con la optimización de los parámetros se considera que la red ya ha aprendido suficiente. Es entonces cuando comienza la tercera fase: la red neuronal se puede usar para su propósito, a la que llamamos fase de inferencia. En este punto, la exigencia computacional es relativamente baja si la comparamos con el entrenamiento, de ahí que una red neuronal ya entrenada pueda desplegarse incluso en un dispositivo como un teléfono móvil. Este es el flujo de trabajo más común para todos los tipos actuales de la IA.

En resumen, ¿cómo una IA empezó a aprender de los humanos?

- El paradigma de una IA basada en datos, que se fundamenta en el uso de redes neuronales artificiales, consiste en proporcionar a la IA muchos datos junto a las herramientas necesarias para poder aprender de ellos.
- Con redes neuronales el programador escribe un código que le permite al algoritmo aprender e incorporar el conocimiento generado por sí mismo a partir de los datos aportados por los humanos.
- Las redes neuronales se popularizaron cuando en 2012 un equipo de investigación de la Universidad de Toronto participó en una competición de visión por computador llamada ImageNet, y la ganó con una red neuronal ejecutada en GPU.
- El flujo de trabajo de definir una arquitectura para crear una red neuronal, entrenarla y finalmente usarla para inferencia es común para todos los tipos actuales de IA.
- El paradigma de una IA basada en datos permite manejar mejor la incertidumbre del mundo real que el paradigma de la IA basada en el conocimiento, al permitir soluciones menos centradas en la ingeniería.
- En 2016, el juego de go fue conquistado por una IA llamada AlphaGo, que venció al campeón mundial de go, Lee Sedol, que utilizaba principalmente el paradigma de una IA basada en datos.

4.
¿Cómo una IA consiguió aprender por sí misma?

Lógicamente, pronto surgió la idea de desarrollar una IA capaz de aprender por sí misma sin necesidad de disponer de datos. Este enfoque dio lugar a una tendencia de investigación en el aprendizaje por refuerzo, esto es, aprender mediante la experiencia.

La IA basada en la experiencia

Aprender interactuando con nuestro entorno es lo primero que nos viene a la mente cuando pensamos en cómo aprende un bebé. Estas interacciones son, sin duda, una fuente esencial de conocimiento sobre nuestro entorno y sobre nosotros mismos, por eso están en la base de casi todas las teorías del aprendizaje. Por ejemplo, cuando aprendemos a conducir, somos plenamente conscientes de cómo responde el entorno a lo que hacemos, así como también buscamos influir en lo que sucede a través de nuestras acciones.

El enfoque del aprendizaje por refuerzo (*reinforcement learning,* en inglés) consiste en aprender de manera dirigida a objetivos mediante la interacción. No introducimos en la IA qué *hacer,* sino qué ha de *descubrir* por sí misma mediante su experiencia de «prueba y error». La IA debe analizar qué produce la mayor recompensa y, por tanto, decidirse por estas opciones y actuar en consecuencia en el futuro. Ese es en realidad

su objetivo: maximizar la recompensa.

En el caso del ajedrez, por ejemplo, quien quisiera aprender a jugarlo de manera autodidacta debería adoptar el rol de los dos jugadores, así, cada vez que ganara o perdiera, aprendería de sus errores y aciertos en cada uno de los roles y ajustaría su estrategia para ganar en las próximas partidas. Eso es lo que los científicos se propusieron conseguir: una IA que aprendiera por sí misma a ganar en un juego de mesa compitiendo contra sí misma, sin la guía de expertos en el juego. Y esta aproximación la probaron con tres juegos de mesa: ajedrez, go y shogi, el equivalente japonés del ajedrez.

Para lograrlo, se diseñó un novedoso sistema que comienza con una sola red neuronal que no sabe absolutamente nada sobre el juego —excepto las reglas— y compite contra sí misma. El rendimiento del sistema mejora con cada iteración y la calidad del juego automático avanza, las redes neuronales son cada vez más precisas hasta lograr el objetivo: ser el mejor jugador del mundo.

Si bien en este enfoque la IA ya no necesita datos para que su algoritmo aprenda, sí requiere muchísima computación, con más potencia, nuevos métodos y tecnología, para poder recrear esta infinidad de partidas. Todo un reto. Dado que el paradigma de aprender de la experiencia (jugando consigo mismo muchas veces) es mucho más costoso a nivel computacional, la solución consiste en utilizar chips aceleradores en paralelo en un mismo ordenador y, al mismo tiempo, disponer de varios ordenadores que trabajen de forma simultánea, con técnicas de paralelismo (permitir que muchos chips colaboren en paralelo para acelerar los cálculos) ya habituales en el mundo de la supercomputación.

Cuando una IA es capaz de aprender por ella misma a jugar

Fueron los mismos investigadores de DeepMind los que se propusieron

este reto. Tomaron una IA de propósito general (es decir, utilizaba la misma estrategia de aprendizaje tanto para el ajedrez como para el go) que pudiera aprender desde cero por sí sola y alcanzar un nivel de juego superior a cualquiera, tan solo conociendo las reglas del juego y sin ningún conocimiento previo de las experiencias de los mejores jugadores.

Para esta nueva IA se diseñó una red neuronal que jugase millones de partidas contra sí misma en un proceso de prueba y error. De esta manera, tomaba nota del tipo de jugadas que contribuían al objetivo de ganar la partida. En este caso concreto, se permitió que la IA refinara la técnica, así que, en lugar de analizar todos los movimientos posibles, el sistema se centró únicamente en aquellos más prometedores según su experiencia previa.

Esto es lo que en 2018 se dio a conocer públicamente, cuando la revista *Science* publicó que se había creado un algoritmo que permitía a una IA aprender por sí misma. La llamaron AlphaZero y su objetivo era vencer al mejor jugador mundial de cada uno de los juegos, que ya no eran humanos: AlphaGo para el go y el programa informático Stockfish para el ajedrez (el programa con el que entrenan los grandes maestros para mejorar su estrategia).

Esta vez no se utilizaron GPU sino un sistema supercomputador basado en un nuevo chip acelerador propio llamado Tensor Processing Unit (TPU), construido por Google, y que solo unas pocas corporaciones con grandes presupuestos podían permitirse.

Un detalle interesante es que dado que AlphaZero no aprendió imitando a grandes maestros, sino a base de prueba y error, desarrolló un estilo propio de juego único y creativo, y mostró nuevas técnicas inéditas al conocimiento humano hasta aquel momento.

Además del objetivo de crear al mejor jugador, los ingenieros del equipo de AlphaZero tenían desde un principio otras ambiciones: «El objetivo de DeepMind es construir sistemas que puedan solucionar algunos de los problemas más complejos del mundo y crear un programa

que puede enseñarse a sí mismo cómo dominar el ajedrez y el go desde cero, es un importante primer paso en ese camino».[3]

La IA, útil más allá de los juegos

Pudiendo aplicar la IA a casi cualquier ámbito, ¿por qué este empeño en centrarse en los juegos? Oriol Vinyals (director de investigación de DeepMind) argumenta que los juegos son muy útiles para la investigación porque ofrecen un entorno controlado en el que hacer pruebas y en el que, además, es muy sencillo definir los objetivos. Asimismo, añade, ayudan a acelerar los resultados. Puedes ejecutar 1.000 juegos en paralelo sin el gasto que supondría, por ejemplo, reproducir 1.000 experimentos en un laboratorio. El propósito final, no obstante, no es encontrar soluciones vencedoras para los juegos, sino utilizar las técnicas aprendidas para solucionar problemas reales.

Veamos un ejemplo. A finales de 2020, la empresa DeepMind anunció un avance —inspirado en el mismo tipo de algoritmo que utilizaba AlphaZero— que ha marcado un punto de inflexión en el campo de la biología computacional. En concreto, logró utilizar este tipo de algoritmos para predecir cómo se plegará una molécula de proteína en su forma final en función del código genético. Este nuevo *software* fue bautizado como AlphaFold. Curiosamente, no ha tenido tanta repercusión en los medios en comparación con los resultados de las IA en los juegos.

Recordemos que el ADN almacena la información genética que determina la secuencia de aminoácidos necesarios para fabricar proteínas, las responsables de casi cualquier proceso biológico. Aunque conocer la secuencia de aminoácidos de una proteína es algo muy importante, la función que una proteína realiza depende de su estructura en tres dimensiones, por lo tanto, conocerla es clave para investigar su función y las implicaciones que pueda tener en cada

proceso bioquímico. Si una proteína se pliega incorrectamente deja de desarrollar la función biológica que tiene asignada, por lo que conocer la relación entre las proteínas y la enfermedad es importante para desarrollar tratamientos.

El proceso mediante el cual una secuencia de aminoácidos llega a su estructura en el espacio de tres dimensiones no es fácil de describir. De hecho, se trataba de uno de los grandes retos de la ciencia, dado que el número de formas posibles es prácticamente infinito y no hay nada en las secuencias de ADN ni de aminoácidos resultantes que permita predecir su forma final. Y es justamente su forma lo que determina su función.

Se puede tardar años en calcular todas las formas posibles de una sola proteína con los métodos convencionales, lo que ha requerido una gran cantidad de horas de cálculo de supercomputadores. Sin embargo, AlphaFold ha conseguido realizar predicciones de las estructuras tridimensionales de casi todas las proteínas conocidas a partir de su secuencia de aminoácidos. Además, se han catalogado para ofrecerlas de forma abierta. Esto permite una enorme variedad de avances, desde el diseño de fármacos completamente nuevos a una mejor comprensión de los modos en que las proteínas pueden plegarse de forma incorrecta, un fenómeno asociado a muchas enfermedades. Sin duda, el logro alcanzado por AlphaFold marca el advenimiento de una nueva era de innovación médica y farmacéutica.

No en vano, el XV Premio Fundación BBVA Fronteras del Conocimiento en Biomedicina ha sido concedido en la edición de 2023 a David Baker, Demis Hassabis y John Jumper «por sus contribuciones al uso de la Inteligencia Artificial para la predicción exacta de la estructura tridimensional de las proteínas».

Ejemplos como AlphaFold muestran cómo se pueden usar las redes neuronales y el aprendizaje por refuerzo para resolver problemas reales. A medida que la IA se aplica con éxito en más y más áreas se hace evidente que se está convirtiendo en una tecnología imparable y de

consecuencias únicas.

¿Qué tipo de IA llegaremos a conseguir con esta tecnología? Aún no lo sabemos, pero en todo caso ha permitido un crecimiento acelerado de los avances algorítmicos en los últimos años. La primera aproximación que tuve al estudio del aprendizaje por refuerzo fue en Cádiz, durante la Machine Learning Summer School en 2016. Oriol Vinyals, en aquel entonces investigador de Google, me aconsejó asistir para ponerme al día de lo que se estaba «cociendo» en el área.

Uno de los cursos a los que asistí fue sobre Deep Reinforcement Learning, impartido por John Schulman (uno de los cofundadores de OpenAI y, en aquel momento, doctorando en la Universidad Berkeley). Aquella charla fascinante y tremendamente difícil de seguir, he de confesar, despertó en mí el interés por el tema, en particular su conexión con la supercomputación (posteriormente ocupé mis horas de confinamiento durante la pandemia en escribir un libro de contenido técnico sobre el aprendizaje por refuerzo).

¿Por qué decía antes que había propiciado un crecimiento acelerado? Seguramente, quizá solo una parte de los asistentes a aquel curso conociera esta tecnología, y solo dos años más tarde DeepMind demostró al mundo entero que el aprendizaje por refuerzo permite que las máquinas aprendan solas. Aquel joven John Schulman comenzó a utilizar en OpenAI el aprendizaje por refuerzo para dirigir el equipo que perfecionó los modelos de lenguaje como GPT (Generative Pre-trained Transformer), que es el motor que mueve ChatGPT y que en solo seis años ha sido capaz de redactar y responder al nivel de un humano.

En resumen, ¿cómo una IA consiguió aprender por sí misma?

- El aprendizaje por refuerzo es una estrategia por la que el aprendizaje de una IA, habitualmente basada en redes neuronales, el conocimiento se consigue de la experiencia.
- En concreto, a la IA que aprende no se le dice qué acciones realizar,

sino qué debe descubrir por sí misma, mediante «prueba y error», y qué acciones producen la mayor recompensa y por tanto decidirse por estas.

- En el caso del ajedrez o el go, AlphaZero consiguió ser el mejor jugador. Como la IA aprendió desde cero, sin riesgo de adquirir hábitos de los grandes maestros, desarrolló su propio estilo de juego único y creativo.
- Si bien en este enfoque a la IA ya no le hacen falta datos, para compensar, lo que sí necesita son supercomputadores para recrear la infinidad de partidas necesarias.
- Un ejemplo de las muchas utilidades de este paradigma basado en la experiencia es AlphaFold, que logró predecir cómo se plegará una proteína en su forma final tridimensional en función del código genético.

5.
¿Cómo una IA ha conseguido ser creativa?

El debate sobre la «creatividad» de la IA está sobre la mesa desde que se popularizaron las primeras IA capaces de generar sorprendentes imágenes a partir de un simple texto. Para muchos, la creatividad es una cualidad intrínsicamente humana que va más allá de las artes y las ciencias, como lo demostró Ferran Adrià en la gastronomía o Johan Cruyff en el fútbol. Es decir, entendemos la «creatividad» como algo más que una creación, sino también como algo innovador que transforma e inspira a otros a continuar siendo creativos.

¿Podemos hoy en día considerar creativa una IA? Está claro que su aplicación en el arte cambiará de manera contundente la naturaleza del proceso creativo. Es más, dejará de ser una herramienta de ayuda a la creación para convertirse en una compañera e, incluso, una agente creativa en sí misma.

Aunque hay quienes no dudan en considerar creativas a las IA como ChatGPT o DALL-E, otras voces, en cambio, opinan que no pueden ser tratadas de «creativas», puesto que tan solo «generan» respuestas basadas en el aprendizaje propiciado con millones de datos creados, en última instancia, por humanos. Es por eso por lo que solo se les concede la etiqueta de «IA generativas».

IA generativa

En 2021, OpenAI lanzó una nueva IA para crear imágenes a partir de texto, bautizada como DALL-E, en un guiño al pintor Salvador Dalí. Este nuevo *software* había aprendido a partir de una gigantesca base de datos con millones de imágenes descritas en texto. Un año más tarde, la organización lanzó de manera gratuita ChatGPT, el archiconocido bot conversacional que emula la lógica del pensamiento humano en su forma comunicativa.

Los avances de las IA están conquistando ámbitos humanos, como el lenguaje o la creación artística de imágenes, mucho más rápido de lo esperado. En los últimos años hemos visto cómo las grandes compañías tecnológicas han volcado gran cantidad de recursos en el desarrollo de productos en esta dirección y están reorientando su estrategia, lanzando continuamente nuevas versiones de sus productos con capacidades cada vez superiores, que ya no incluyen solo textos de mayor calidad sino que incorporan una IA multimodal, es decir, capaces de trabajar con texto, imágenes y sonido a la vez.

La fascinación que despertó ChatGPT fue tal que en tan solo dos meses desde su lanzamiento, a finales de 2022, sumaba ya 100 millones de usuarios únicos. Para hacerse una idea de la magnitud de lo que representa, Instagram tardó más de dos años en llegar a ese mismo número de usuarios. Estas cifras han convertido ChatGPT en la aplicación de consumo en Internet de más rápido crecimiento de la historia hasta el momento. Quizá la fascinación se debió a que, para nosotros, los humanos, el lenguaje es una ventana a la inteligencia. De cualquier forma, sea bienvenido este interés repentino por la IA de parte de la sociedad.

Cómo se llega a un bot conversacional que parece un humano

Todo empezó con la fundación en 2015 de la empresa OpenAI por parte

de un grupo de líderes tecnológicos (Elon Musk entre ellos, aunque tres años después abandonó el proyecto con el pretexto de que la empresa no estaba investigando según los objetivos fundacionales, y el tiempo parece que le ha dado la razón; ahora es Microsoft la compañía que más está invirtiendo en ella). Desde entonces, OpenAI se ha centrado en construir redes neuronales cada vez de mayor tamaño, entrenadas a su vez en supercomputadores cada vez más potentes, a los que conocemos como supercomputadores a gran escala.

La primera versión de red neuronal que presentó OpenAI fue el modelo de lenguaje GPT, lanzado en 2018, que utiliza 117 millones de parámetros. Los modelos de lenguaje son un tipo de red neuronal entrenados con enormes cantidades de secuencias de letras y palabras de diferentes longitudes, y utilizan un mecanismo diferente al resto de redes neuronales: los GPT están diseñados para prestar atención a distintas partes de una frase con el fin de crear relaciones entre ellas. De esta manera, rastrea dónde aparece cada palabra o frase dentro de una secuencia, gracias a lo cual puede *interpretar* el significado de las palabras según el contexto. Estas probabilidades se calculan a partir de una selección de muchísimos textos en los que el programa busca con qué otras palabras se asocia más frecuentemente cada palabra.

En 2019, OpenaAI presentó la siguiente versión de este potente sistema de lenguaje natural, el GPT-2, compuesto ya por 1.500 millones de parámetros. Esta nueva versión estaba ya configurada de forma tal que con una breve indicación escrita (de solo una o dos frases) fuera capaz de generar una narración completa.

En mayo de 2020 llegó GPT-3, un sistema cien veces más potente que el anterior, con más 175.000 millones de parámetros. El cambio fue sustancial. Los diferentes GPT se entrenan para que sean capaces de adivinar cuál debe ser la siguiente palabra de una frase. Es decir, el modelo genera un texto palabra a palabra, ejecutándose iterativamente el algoritmo de predicción una y otra vez para cada nueva palabra. La red neuronal GPT-3 se entrenó con miles de millones de textos de

diferentes fuentes de Internet, desde libros y páginas web a conversaciones reales entre usuarios. Para que el lector se haga una idea de la dimensión: la Wikipedia entera constituye alrededor del 3 % del total de información con que se *alimentó* al nuevo programa.

El entrenamiento es el siguiente: se oculta una palabra del texto y se ejecuta la red neuronal para que la prediga. De esta manera, el esquema de entrenamiento es equivalente al del ejemplo de la red neuronal que clasificaba imágenes según aparecía un gato, o no, en ella. En este caso, sabemos cuál es la solución que buscamos porque es precisamente la palabra que se le ha ocultado al programa. Finalmente, se compara el valor que ha calculado la red neuronal con el esperado para ajustar los valores de los parámetros de la red neuronal (ver capítulo 3).

Recordemos que una interpretación simple de lo que contienen los miles de millones de parámetros de la red neuronal es su versión comprimida de todo el conocimiento que se les ha mostrado para aprender. Es un proceso similar al de comprimir un archivo. Requiere dos pasos: primero, la codificación para comprimir, durante la cual el archivo se convierte a un formato más compacto, y luego la decodificación a partir de la información comprimida, en la que se invierte el proceso. Es decir, cuando usamos estas IA para generar texto, en realidad estamos decodificando y, por tanto, la secuencia exacta de palabras que estaban en los datos originales no se encuentra almacenada tal cual, puesto que la copia comprimida solo es una representación de la información real. Sin embargo, al decodificar, es posible obtener una aproximación en forma de texto gramatical equivalente. Esto explica algunos casos en los que las respuestas de las IA son poco acertadas, pues en cierta manera es inevitable que se haya perdido información en el proceso de *compresión*.

Es importante remarcar que estas IA generativas del estilo de GPT tienden a ser adaptables, lo que significa que pueden adquirir otras habilidades aparte de aquellas para las que fueron explícitamente capacitadas. Esto es posible gracias a su entrenamiento generalista.

GPT-3, por ejemplo, no solo aprendió a escribir un texto de aspecto realista, sino que también aprendió a generar un código de programación aceptable, a pesar de que no tenía la intención explícita de hacerlo al principio.

A finales de 2022 se lanzó la versión abierta de ChatGPT (una versión mejorada de GPT-3). En esta ocasión, el programa se centró en utilizar el contenido de las conversaciones interactivas entre personas. Recordemos que el modelo del lenguaje GPT-3 solo estaba entrenado para predecir la siguiente palabra en una secuencia de texto, pero era incapaz de *comprender* su significado. En la nueva versión se mejoró el proceso de aprendizaje con la inclusión de comentarios humanos y con técnicas de aprendizaje por refuerzo, pero usando una retroalimentación con intervención humana en el ciclo de entrenamiento. ¿Por qué? Porque basar el entrenamiento en textos extraídos de Internet había tenido un efecto colateral indeseado: junto con la información válida, GPT-3 había absorbido gran parte de la desinformación y sesgos que se encuentran en la red. Por ello, para reducir la cantidad de información errónea y textos ofensivos que producía GPT-3, hubo que ajustarla de forma «manual». Al final, el proceso de entrenamiento de las IA generativas requiere que la mano humana esté muy presente.

Estos modelos siguen evolucionando a medida que escribimos su historia. Poco después de lanzar el primer ChatGPT, OpenAI anunciaba una nueva versión, GPT4, que admite como entrada no solo texto sino también imágenes. En esta ocasión, la compañía no ha hecho público ni su entrenamiento ni detalles sobre los parámetros o requerimientos computacionales. En todo caso, representa una mejora de la versión anterior, pese a que aún es propenso a los mismos tipos de problemas de veracidad.

Poder transformacional de las IA generativas

ChatGPT todavía no es capaz de leer un libro (entendiendo por leer la facultad de comprender su contenido). La aproximación actual de cualquiera de las IA disponibles se basa en la representación de probabilidades intentando adivinar qué palabras tienden a concurrir en una frase o contexto. Es decir, se generan textos que parecen escritos por humanos, pero no significa que la IA tenga conocimiento del tema ni que haya comprendido el texto. Es algo similar al predictor de texto de WhatsApp, que nos sugiere palabras para completar el mensaje que estamos escribiendo.

Uno de los aspectos más preocupantes es la falta de veracidad y los sesgos que mencionábamos en el apartado anterior. En muchas ocasiones la IA es capaz de responder con información falsa como si fuera cierta, bien sea porque los datos de entrenamiento no están actualizados o porque en el proceso de codificación y decodificación se ha perdido información.

No hay que perder de vista que la red neuronal no es más que un modelo limitado del mundo conformado con los valores de sus parámetros, y no un modelo del mundo real. De momento, no se ha encontrado la manera de entrenar modelos con datos extraídos de Internet sin absorber lo que se conoce como la suciedad de los datos, es decir, los bulos y contenidos ofensivos, entre otros. De ahí que, igual que hemos visto en el caso de GPT-3, la única solución que existe hasta el momento es que operadores humanos filtren la información a mano.

A pesar de todas estas limitaciones, la ola de las IA generativas no ha hecho más que empezar, y ChatGPT representa solo el primer exponente de los nuevos modelos que pronto estarán presentes en todos los aspectos de nuestras vidas. Sus posibilidades son casi infinitas: desde chatear a generar documentos sofisticados o sencillamente servir de inspiración. Como ya ha ocurrido con los traductores automáticos, las aplicaciones de creación de texto serán habituales en nuestros dispositivos y pasarán a integrarse en el *software* que utilizamos cotidianamente en nuestras rutinas productivas, tanto en el ámbito

doméstico como en el empresarial y educativo.

Los gigantes tecnológicos habituales se han lanzado a una carrera para desarrollar IA generativas no solo mucho más potentes y eficientes, sino también específicas para ámbitos determinados, es decir, entrenadas con datos personalizados (por ejemplo, medicinal o empresarial). Pronto se convertirán en las mayores expertas en el área en la que hayan sido entrenadas.

Un último apunte sobre esta carrera por la IA: la proliferación de desarrolladores ha generado también tensiones entre la comunidad de código abierto IA y las empresas privadas respecto a la exclusividad de los códigos de IA. Los primeros abogan por unas IA generativas como instrumento de creatividad e innovación abiertas, mientras que los segundos defienden su privatización.

La IA basada en fuerza bruta

¿Pero cómo hemos llegado aquí? Ya hemos dicho que el año 2012 fue un punto de inflexión para la adopción de las redes neuronales, en especial gracias al equipo de la Universidad de Toronto y su revolucionaria participación en la competición de ImageNet.

Pronto se vio que los métodos de aprendizaje de las redes neuronales podían aprovechar magníficamente técnicas de paralelismo, es decir, utilizar varios chips aceleradores de forma simultánea para reducir el tiempo de entrenamiento de las redes neuronales. Y, por último, los supercomputadores a gran escala permitieron acelerar esto aún más, interconectando una gran cantidad de máquinas con varios chips aceleradores cada una.

Sin duda, el paralelismo es una técnica capital en la supercomputación a gran escala. Pongamos como ejemplo la red neuronal de Google para traducción multilingüe. Consiste en una red con 600 mil millones de parámetros, es decir, una capacidad de

computación equivalente a 22 años si solo se dispusiera de un chip acelerador tipo TPU (como el que se usó en AlphaZero). Pero dado que el sistema de Google utiliza 2.048 chips de este tipo en simultáneo, consigue realizar el entrenamiento en solo cuatro días.

En el último año, las necesidades de computación para entrenar las IA generadoras de texto se han multiplicado por dos cada tres o cuatro meses, con lo que las infraestructuras con gran capacidad de computación se han revelado como fundamentales. Hoy en día es inconcebible pensar en un supercomputador a gran escala que no cuente con un *hardware* pensado para entrenar una IA. Uno de los más recientes es el MareNostrum 5 (ha entrado en funcionamiento durante 2023), que incluye 4.480 chips aceleradores GPU de última generación fabricados por Nvidia. Es uno de los nodos principales de la red europea de supercomputación EuroHPC.

La gran capacidad computacional disponible en la actualidad ha permitido a la comunidad de IA avanzar los últimos años con mucha rapidez y diseñar redes neuronales cada vez más complejas, aunque esto ha exigido aumentar la infraestructura de computación a niveles nunca vistos. Estamos inmersos en lo que podríamos llamar el paradigma de la IA basada en «fuerza bruta». Es decir, algoritmos de miles de millones de parámetros que necesitan supercomputadores a gran escala para ser entrenados con ingentes cantidades de datos. Unos recursos, desde luego, solo al alcance de muy pocos.

En resumen, ¿cómo una IA ha conseguido ser creativa?

- ChatGPT fue una IA generativa orientada a la comprensión y generación de diálogos para conversar con los humanos basada en un tipo de red neuronal de miles de millones de parámetros.
- Este tipo de redes puede adquirir habilidades mucho más complejas de aquellas para las que fueron capacitadas debido a su entrenamiento generalista con grandes conjuntos de datos.

- Para evitar absorber en lo posible la desinformación y los sesgos que contiene Internet, se ha mejorado el entrenamiento de las redes neuronales con técnicas de aprendizaje por refuerzo en el que intervienen equipos humanos.
- La ola de IA generativas no ha hecho más que empezar. ChatGPT solo representa el primer exponente de unas IA generativas que estarán pronto presentes en todos los aspectos de nuestras vidas y conllevará un impacto social sin precedentes.
- La IA generativa está basada en el paradigma de fuerza bruta, un escenario en el que redes neuronales de miles de millones de parámetros deben ser entrenadas con grandes cantidades de datos, para lo que se requieren supercomputadores a gran escala.

6.
¿Podrá una IA llegar a pensar?

Todas las IA de los paradigmas que hemos presentado hasta ahora funcionan para propósitos específicos y para realizar tareas que antes solo podían hacer las personas. La pregunta que ahora se plantea la comunidad científica es: ¿suponen estas IA un avance importante hacia una inteligencia artificial general, parecida a la humana? Evidentemente, no estamos hablando de una inteligencia orgánica, pero ¿puede llegar a ser comparable a la humana? ¿Podemos hablar de máquinas que consigan «pensar»?

Falta de sentido común de la IA

Aunque las IA aún están lejos de ser como las inteligencias humanas, algunos investigadores consideran que ejemplos de IA como ChatGPT, creadas a base de agregar billones y billones de datos, son la antesala a sistemas que razonen como los humanos. Es lo que se denomina IA General —en la literatura académica se conoce también como IA Fuerte —. Hasta el momento, solo se ha conseguido que las IA puedan resolver problemas concretos como redactar un texto. Es lo que llamamos IA Específica —o IA Débil o Estrecha, en la literatura—. No es poca cosa: estas IA son tan buenas como un humano a la hora de resolver un problema concreto.

De momento no sabemos cómo conseguir una IA que pueda realizar

tareas para las que no se las haya entrenado específicamente o para las que se requiere de improvisación. Una de las limitaciones más claras que presentan las IA en su formato actual es que, pese a su capacidad para hacer cálculos complejísimos en milisegundos, carecen de un elemento difícil de parametrizar en términos científicos: el sentido común.

Veamos el caso de los coches sin conductor. Hace años que se pronostica su implantación en la sociedad, pero lo cierto es que no acaba de llegar. Uno de los mayores desafíos por resolver consiste en ajustar la IA que gobierna el sistema de conducción no tripulada para que tenga una mejor capacidad de reacción y resolución de eventos inesperados. Aunque en la codificación y el entrenamiento de las redes neuronales se tienen en cuenta un gran número de variables, siempre hay imprevistos que hacen fallar al sistema porque este carece de la capacidad de improvisación y de sentido común del ser humano. En términos generales, la circulación autónoma está muy avanzada en entornos previsibles (como autopistas o zonas residenciales con poca densidad de población y tráfico), pero la situación es distinta cuando hablamos de una ciudad de mayores dimensiones. En estos casos, el problema son las interacciones inusuales con el entorno, como peatones que no respetan el semáforo y cruzan las calles en rojo o en zonas prohibidas, o patinetes que aparecen de la nada y son difíciles o imposibles de interpretar con precisión para un coche autónomo. Normalmente no nos detenemos a pensar en ello, pero la coordinación entre conductores y peatones depende de detalles tan sutiles como el contacto visual o el movimiento de una mano, algo que todavía es muy complicado de reproducir o procesar para un coche autónomo.

En resumen, el sentido común (el bagaje de factores psicológicos, biológicos y experienciales que nos permite proceder como lo haría la mayoría de las personas en circunstancias similares) es lo que nos permite tomar decisiones en entornos de incertidumbre, algo que por ahora no puede hacer ninguna IA, dado que modelar matemáticamente la incertidumbre es muy complejo. La IA actual es eficaz en tareas

específicas, pero presenta dificultades para adaptarse a situaciones inesperadas a las que no puede comprender ni aplicar conceptos abstractos. Dotar a la IA de sentido común es una cuestión crucial si queremos llegar a una IA General, pero ¿cómo conseguirlo? Por el momento los investigadores en IA no hemos visto ningún indicio del camino a seguir para resolver este problema porque, como hemos visto a lo largo de estos capítulos, las redes neuronales necesitan muchos más datos de entrenamiento que un niño para aprender.

El sistema AlphaZero aprendió por sí mismo a jugar al ajedrez, pero si cambiásemos el tamaño del tablero, por ejemplo, no sería capaz de jugar, puesto que ya no sería el mismo juego del que tiene referentes sino uno totalmente nuevo, y AlphaZero debería volver a ser entrenada. En cambio, los seres humanos sí sabríamos adaptarnos al juego. Y, además, cuando la IA aprenda a jugar con el nuevo tablero, olvidaría cómo se hace con el tablero anterior. Esto es lo que se conoce como «olvido catastrófico»: cuando le introducimos un nuevo conocimiento las IA basadas en redes neuronales olvidan todo lo que habían aprendido hasta el momento.

Esto se debe a que las redes neuronales no están preparadas para entender por qué las cosas suceden, solo saben encontrar patrones y conexiones entre esas cosas. Esto significa que, aunque puedan predecir muy bien lo que va a pasar en una situación dada, no son capaces de explicar por qué va a suceder. Es decir, no pueden proporcionar una explicación causal de por qué se produce ese resultado. Esta es una limitación importante cuando se necesita entender exactamente por qué ocurre algo. El matemático y filósofo Judea Pearl lo ejemplifica muy bien con dos elementos: la salida del sol y el canto del gallo. Una IA es capaz de comprender la correlación entre estos dos hechos, pero no es capaz de identificar si el gallo canta porque sale el sol, o si es el sol el que sale porque canta el gallo.

Se requieren nuevos enfoques para el avance algorítmico

Actualmente nos encontramos con IA Específicas que son capaces de hacer muy bien —a menudo mejor que el humano más capacitado— una tarea concreta, por ejemplo, jugar al ajedrez. Pero ni siquiera las IA más avanzadas son capaces de pensar. Lo que hacen, como hemos visto, es crear una copia comprimida (almacenada en los parámetros de su red neuronal) de toda la información que han recogido durante la fase de aprendizaje. Estas IA se autoajustan (modificando el valor de sus parámetros) a medida que aprenden, con lo que llegan a generar frases coherentes. Una vez entrenadas, nadie (ni siquiera sus creadores) sabe cómo se organiza internamente (cómo se almacena el conocimiento en sus parámetros), pero emerge la capacidad lingüística. Inevitablemente, en esta copia comprimida hay una pérdida de información. Por ello, aunque en la mayoría de los casos estas IA muestran resultados «inteligentes» cuando se escalan a niveles gigantescos, aún tienen lugar situaciones en las que son incapaces de producir la información esperada. Es en esos momentos cuando nos damos cuenta de que no «piensan».

En realidad, que una IA pudiera pensar significaría que podría decidir su propósito y no limitarse a hacer lo que el humano le indica. En cualquiera de los paradigmas de las IA presentados en este libro es necesario que un ser humano pulse el botón de «inicio» para una tarea específica. Cuando se hipotetiza sobre una IA General se hace referencia a una IA «consciente», capaz de reflexionar y reconocerse en el mundo en el que se desenvuelve, aunque hablar de «conciencia artificial» es un concepto harto complicado. Podríamos resumir, entonces, que la IA General aspira a desarrollar sistemas IA «generalistas».

A menudo, especialmente en la ciencia ficción, se habla de superinteligencia artificial, término algo controvertido y poco claro, para referirse al desarrollo de sistemas con inteligencia superior a la humana. Por lo general, el concepto siempre tiene algo de profecía sobre

el advenimiento de la singularidad tecnológica, el momento en que las máquinas se vuelven más inteligentes que los seres humanos.

Más allá de la ciencia ficción, la ciencia también muestra sus reticencias al desarrollo de las IA. Nick Bostrom, director del Instituto del Futuro de la Humanidad de Oxford, analiza en su libro *Superinteligencia*[4] los riesgos para la humanidad de las IA superinteligentes. La tesis que defiende es que, si una IA llegara a ser más inteligente que nosotros, tendría capacidad de automejorarse y no seríamos los humanos quienes dirigiríamos sus avances sino ella misma y, como consecuencia, en algún momento nos superaría en capacidades. Un argumento recurrente postula que una IA, con el objetivo de preservar el planeta, podría llegar a la conclusión de que somos los humanos los que estamos destruyendo el medio ambiente —¿a quién no se le ha pasado por la cabeza alguna vez?— y decidir que debemos ser eliminados.

Quienes creen en la singularidad argumentan que los avances exponenciales en IA la hacen inevitable, aunque muchos científicos somos escépticos respecto a la llegada de la singularidad. La principal razón es que esta hipótesis se basa en la continuación de un progreso exponencial y esto, claramente, no es realista. Con ello no quiero decir que no haya motivos de preocupación por el tema, sino simplemente que no vemos que sea algo imperturbable ni inminente.

No sabemos cuán lejos o cerca estamos de la IA General. En todo caso, conocemos las diferencias entre los mecanismos de aprendizaje de las IA y los sistemas inteligentes biológicos, y esto nos puede dar una pista del camino que todavía queda por recorrer.

Si tomamos como ejemplo el cerebro humano —aunque no es el único con el que se compara—, una de las diferencias más significativas se encuentra en que los sistemas de aprendizaje biológicos suelen tener suficiente con unas pocas observaciones o experiencias para aprender algo nuevo, y lo hacen además de forma incremental y constante a medida que interaccionan con el entorno. En cambio, los actuales

sistemas de IA necesitan una gran cantidad de datos y, como ya hemos visto, su conocimiento no es flexible.

De hecho, la complejidad de las neuronas biológicas y sus conexiones dista mucho de las de las neuronas artificiales. El cerebro humano activa un subgrupo concreto de neuronas para procesar la información, que tienen plasticidad y que constantemente se están regenerando y creando nuevas interconexiones. De momento, las redes neuronales artificiales están lejos de tener esta capacidad. En palabras del Dr. Ramon López de Mántaras (fundador del Instituto de Investigación en Inteligencia Artificial del CSIC), «hoy por hoy no existe ninguna IA capaz de contextualizar y de hacer el tipo de inferencias básicas que incluso un niño realiza sin esfuerzo».

Basándose en estas diferencias evidentes, muchos investigadores de la IA sugieren que se debe abordar el problema con un enfoque totalmente diferente al actual, un cambio de paradigma que debería incluir, como mínimo, conceptos tan humanos como la abstracción o la causalidad.

La IA es un problema de supercomputación

Ahora mismo no hay líneas de investigación suficientemente maduras para sembrar los cimientos del próximo paradigma en IA (y que nos acerque más a la IA General). Por eso, solo sabemos seguir exprimiendo el paradigma de una IA basada en fuerza bruta, logrando mejoras continuas en la frontera de la computación que utilizamos, del volumen de los datos que le suministramos y del tamaño de la red neuronal que construimos.

En realidad, este es el camino que está llevando OpenAI con su GPT. El rendimiento de GPT-4 es claramente mejor que el de GPT-3 y, según un estudio de marzo de 2023 realizado por investigadores de Microsoft, ya hay tareas en las que su rendimiento llega a ser cercano al nivel humano. El estudio, acaso excesivamente entusiasta, llega a afirmar que

«GPT-4 muestra chispas de IA General», algo que, según sus autores, se demuestra por la variedad de áreas en las que ha adquirido experiencia (como la literatura, la medicina y la programación, entre muchas otras) y la diversidad de tareas que puede realizar en cada una de estas áreas. Sin embargo, también es necesario remarcar que muchos investigadores han puesto en duda la fiabilidad de los resultados de este estudio, ya que las características de GPT-4 no son de dominio público.

Es el momento de plantearnos las siguientes preguntas: ¿hasta dónde podemos llegar con este paradigma de fuerza bruta? ¿Todos los vectores impulsores de la IA pueden continuar creciendo exponencialmente como hasta ahora, especialmente la capacidad de computación?

Es cierto que la evolución en computación ha sido el motor de los avances en cuanto a IA. En los años cincuenta del siglo pasado se ambicionó una máquina que ganara al ser humano al ajedrez. Tuvieron que pasar casi cincuenta años para disponer de la capacidad computacional y de algoritmos capaces de crear una IA como Deep Blue. Entrado el siglo XXI, la utilización de los chips aceleradores GPU en paralelo permitió que los algoritmos basados en redes neuronales, que ya se conocían desde hacía décadas, dieran lugar a programas más complejos como AlphaGo.

En este punto se comprendió que la computación podía ser el factor más determinante en el avance de la IA, y las compañías centraron sus esfuerzos en crear supercomputadores dotados de una gran cantidad de chips aceleradores GPU o TPU. Esto permitió entrenar redes neuronales con la aproximación de aprendizaje por refuerzo. Y, como hemos visto, condujo a la creación de una IA que aprendía por sí misma sin la ayuda de los humanos, la AlphaZero, que se convirtió en la mejor jugadora de juegos de mesa en 2018.

Quizá la mejor evidencia de que la computación es el pilar fundamental del avance de la IA es el advenimiento de las IA generativas que comentamos en el capítulo 5, pues estas no fueron posibles hasta la llegada de los supercomputadores a gran escala. Un

ejemplo: para entrenar el modelo de lenguaje GPT-3 se requirió aproximadamente 600.000 veces más capacidad de computación que la que tenía la IA que ganó el concurso ImageNet en 2012. ¿Podemos mantener esta tasa de crecimiento en requerimientos de computación?

Entiendo que las magnitudes computacionales son difíciles de imaginar. He aquí una comparación: para medir el tiempo necesario para realizar las cantidades abismales de cálculos que se requieren para entrenar una IA de tamaño similar a GPT-3, los requerimientos de computación se miden en una unidad llamada petaFLOPS-día. Un petaFLOPS-día equivale a mil billones de cálculos por segundo (10 elevado a 15), que es equivalente a 10 elevado a 20 operaciones por día. A modo de comparación y redondeando (bastante), se puede considerar que un ordenador estándar necesitaría alrededor de un año para computar un petaFLOPS-día. Por lo tanto, con un simple ordenador estándar necesitaríamos varios milenios para alcanzar los 3.640 petaFLOPS-día necesarios para entrenar a un GPT-3.

Uno de los modelos de lenguaje más grande del que conocemos sus características es PaLM (Pathways Language Model), de Google, con 540.000 millones de parámetros. PaLM requiere 10 veces más cálculos que GPT-3. Según la información que dio Google, se han usado para su entrenamiento «6.144 chips TPU durante 1.200 horas y 3.072 chips TPU durante 336 horas», es decir, se han requerido más de dos meses de ejecución en un supercomputador a gran escala con chips aceleradores TPU de última generación.

Otro aspecto a tener en cuenta es el enorme consumo energético que requiere operar los supercomputadores a gran escala. Producir esta energía tiene un costo para el medio ambiente. Según el artículo en el que se detallan las características de PaLM,[5] Google utilizó principalmente energía «limpia» para entrenar el modelo: «El centro de datos de Oklahoma está alimentado sustancialmente por viento y otras fuentes de energía renovables, y operó con un 89 % de energía libre de carbono durante el período de tiempo que el modelo estuvo

entrenando». Lamentablemente, aunque en este caso se haya usado energía «limpia», no se puede afirmar que sea así en general.

Si no se adoptan nuevas estrategias, más allá de seguir añadiendo chips aceleradores, los requerimientos energéticos de la supercomputación a gran escala se volverán inasumibles, al menos si queremos seguir entrenando las redes neuronales al mismo ritmo que hasta ahora.

Ya hay, de hecho, investigaciones en marcha para solucionar este problema. Las más avanzadas en estos momentos, aunque no se espera poder aplicarlas a corto plazo, son en el área de los microprocesadores. Se está trabajando en la creación de chips especializados de redes neuronales, optimizados para acelerar los cálculos matemáticos para su entrenamiento. Empresas como IBM o Intel han realizado inversiones importantes en una nueva clase de chip que se acerca mucho al cerebro humano: los llamados chips «neuromórficos». Con estos circuitos será posible crear la red neuronal en el propio *hardware* a través de miles de neuronas artificiales interconectadas, en lugar de crearla desde el *software* como hasta ahora. Y además de mejorar los tiempos de cálculo son también más eficientes en el consumo de energía.

Más allá de estos avances, hacen falta con urgencia más soluciones a nivel tecnológico si queremos mantener el crecimiento vertiginoso que hemos tenido hasta ahora. Hay quien apunta a la computación cuántica, que puede tener el potencial de cambiarlo todo. Por el momento, el BSC, nodo principal de la red Quantum Spain —que cuenta con trece centros de supercomputación en toda España—, es donde se instalará una infraestructura de supercomputación cuántica conectada con el supercomputador MareNostrum 5 y a la que podrán acceder investigadores y empresas para experimentar con esta tecnología. Según la Dra. Alba Cervera-Lierta (investigadora del BSC y coordinadora de esta red), la mecánica cuántica, en donde entran en juego características como la superposición o el entrelazamiento, abre puertas a trabajar de forma diferente. De hecho, una de las investigaciones más recientes del

centro pretende determinar qué modelos de IA pueden beneficiarse de ella. A pesar de contar con esta infraestructura, la Dra. Alba Cervera-Lierta aclara que en esta área estamos en una fase muy incipiente y que nos queda mucha investigación por delante para saber si podremos desarrollar los QPU, chips con tecnología cuántica.

En resumen, debemos aceptar que por el momento no hay garantías de que la capacidad de computación pueda seguir creciendo exponencialmente, al menos a corto plazo. Y como hemos visto, el progreso de la IA está estrechamente ligado al aumento de la capacidad computacional. Por el momento, este freno nos permite evitar que las nuevas IA escapen lentamente del control de nuestro entendimiento, puesto que cada vez será más difícil para nosotros seguir el ritmo de asimilación de su complejidad. Se preguntará el lector, entonces, si es prudente integrar estas IA en todas las aplicaciones y dispositivos, como se ve que está sucediendo. ¿Deberíamos quizá ralentizar el desarrollo de las mejoras hasta que las comprendamos mejor?

Una parte de reconocidos investigadores en IA se pronunció en una carta abierta titulada «Pause Giant AI Experiments» (Haced una pausa en los grandes experimentos con IA). En ella, solicitan a todos los laboratorios de IA que detengan los entrenamientos de los sistemas más potentes que el GPT-4 durante un tiempo. También, que se incluya en esta pausa a todos los actores clave, y consideran que si tal parada no se pudiera promulgar con rapidez, los gobiernos deberían intervenir para garantizarla. Esta pausa serviría para desarrollar e implementar un conjunto de protocolos de seguridad compartidos para el diseño y desarrollo avanzados de IA, a la vez que se trabaje con los legisladores para crear sistemas de gobierno con capacidad reguladora sobre las IA generativas tan potentes. El objetivo de este grupo no es detener el desarrollo de la IA, sino dar un paso atrás en esta carrera peligrosa que nos puede llevar a modelos de «caja negra» cada vez más grandes con «habilidades emergentes» que se escapen a nuestra comprensión.

El Santo Grial de los investigadores

«Parece probable que, una vez que el método de pensamiento de la máquina haya arrancado, no debería llevarle mucho tiempo sobrepasar nuestras limitadas capacidades [de los humanos]. Por tanto, en algún momento deberíamos esperar que las máquinas tomaran el control».

La frase es del mismísimo Alan Turing y corresponde a un extracto de una conferencia que dio en 1951. Desde entonces, la búsqueda de una verdadera IA General a nivel humano se ha convertido en el Santo Grial para los investigadores. Para muchos científicos, esta búsqueda es una de las grandes preguntas que dan a la ciencia y la investigación su razón de ser, equivalente a preguntarse el origen de la vida.

Hay quien se muestra más optimista en cuanto a los tiempos que se requieren para lograr esta IA General, pero reputados investigadores como Stuart Russell, coautor junto a Peter Norvig del libro de texto universitario sobre IA más importante del mundo,[6] considera que «vamos a necesitar varios Einsteins para que esto suceda».

Mientras escribía este libro he podido contrastar su contenido con muchos colegas académicos. El doctor Rubén Tous (brillante profesor e investigador) me trasladó mientras tomábamos un café una conclusión más que inquietante para nosotros: coincidíamos que si en 2018, cuando nos encontrábamos en la época del «paradigma de la experiencia», nos hubieran dicho que en 2022 dispondríamos de una IA como DALL-E que permitiría crear imágenes de calidad a partir de texto o que podríamos interactuar con un chatbot conversacional como ChatGPT, ambos —y después he visto que coincidimos con muchos otros de nuestros colegas— habríamos dicho que ni pensarlo, que para estas capacidades habría que esperar todavía muchos años.

A pesar de que continúo creyendo que la IA General, en el sentido más estricto de su definición, se encuentra lejana en el tiempo, y que antes deberíamos ocuparnos con urgencia de la IA actual, la rapidez con la que avanzan últimamente los desarrollos me provoca un cierto

desasosiego, agravado por la inquietud que me genera que investigadores de la talla de Oriol Vinyals afirmen que su generación verá una IA que iguale o supere a la del ser humano.

En resumen, ¿podrá la IA llegar a pensar?

- La IA General hace referencia a una inteligencia que puede ser comparable a la humana y permite «pensar» a las máquinas en el sentido amplio del verbo. En cambio, con la IA Específica nos referimos a aquella que permite hacer una tarea específica, incluso mejor que un humano, pero siguiendo instrucciones claras.
- La IA en su estado actual carece de sentido común, una característica fundamental en la inteligencia humana que nos permite reaccionar ante lo imprevisto. Además, la comprensión de las redes neuronales termina en la correlación, no son capaces de comprender la relación causa-efecto.
- A diferencia de las redes neuronales que requieren muchísimos datos para aprender, los humanos tenemos suficiente con unas cuantas observaciones para aprender algo nuevo, además de que podemos aprender de forma incremental.
- Por singularidad tecnológica se entiende aquel momento en que las máquinas se vuelven más inteligentes que los seres humanos y se basa en la suposición de que los avances en IA sigan siendo exponenciales.
- Todos los esfuerzos para avanzar en las capacidades de la IA consisten en seguir exprimiendo el paradigma de una IA basada en la fuerza bruta. Sin embargo, actualmente ya no hay garantías de que podamos continuar creando innovaciones tecnológicas que permitan mantener el ritmo de crecimiento que se ha mantenido hasta ahora.
- El progreso de la IA a largo plazo está estrechamente ligado al aumento de la capacidad computacional, siendo esta la que asume mayor parte del trabajo de obtener conocimiento directamente de las grandes cantidades de datos y no de los conocimientos humanos.
- Hay investigadores del área que consideran que deberíamos ralentizar

las mejoras de estas IA hasta que las comprendamos mejor.

- El sueño de crear una verdadera IA General a nivel humano es el Santo Grial que están buscando los científicos desde los primeros trabajos de Alan Turing.

7.
¿Nos debe preocupar el impacto de la IA actual?

Excluir a los humanos de la ecuación con la aparición de una IA General o superinteligencia no va a ser inminente, pero no quiere decir que podamos relajarnos. El impacto del uso de las IA Específicas ya disponibles no tiene precedentes. Ciertamente, la IA tiene el potencial de crear muchos beneficios económicos y sociales significativos tanto a nivel colectivo como individual, sin olvidar el potencial que representa para la investigación, algo clave para el futuro de la humanidad. La IA puede aplicarse a cualquier ámbito que pueda digitalizarse, ya sean imágenes estáticas o en movimiento, lenguaje natural hablado o escrito en distintos idiomas o sonidos, pero también asume, cada día más, tareas que hasta ahora habían estado reservadas a la esfera humana. Esto la sitúa en el centro de la transformación del mundo laboral, por lo que ha suscitado el miedo a que destruya empleos y ponga en riesgo las formas de subsistencia de la sociedad. No hay duda de que la IA hará parte de nuestro trabajo, de prácticamente todos nuestros trabajos, el día de mañana.

Sin embargo, a quienes conocemos y trabajamos la IA «por dentro» nos preocupa más la confianza excesiva que la sociedad en general muestra sobre sus capacidades, y la confianza con la que, cada vez más, se delegan en ella responsabilidades en la toma de decisiones. La naturaleza propia de la IA actual es propensa a errores y no nos permite

saber ni explicar con exactitud por qué ha tomado una determinada decisión en un momento dado. Y, lo más importante, la IA no está dotada de la concepción del bien y del mal, por lo que hay que ir con cuidado y atención, pues no sabemos para qué fines puede utilizarse.

Las IA son cajas negras

Las IA actuales basadas en redes neuronales son eficaces en muchos entornos en los que están produciendo resultados muy útiles. Ahora bien, también presentan problemas que no se encontraban en las IA basada en el paradigma del conocimiento. Antes hemos mencionado las «cajas negras», la forma en que denominamos a las redes neuronales para las que no tenemos manera de saber cómo ha llegado a una determinada decisión o predicción sobre los datos.

Hagamos una pequeña recapitulación. En el capítulo 3 vimos que una red neuronal almacena el conocimiento adquirido en sus parámetros. La realidad es, no obstante, que no comprendemos exactamente cómo se almacenan esos datos, por lo que no podemos describirlo en detalle, ya que la forma en que se representa el conocimiento dentro de la red surge de manera orgánica y la representación se distribuye de forma automática entre los millones de parámetros de las redes neuronales.

Esto dificulta la comprensión de por qué una red neuronal ha tomado una determinada decisión, a diferencia de los algoritmos de antaño basados en la regla del *if-then* con los que se podía seguir el camino por el que se había llegado a una determinada decisión. Esta opacidad de las redes neuronales es una de las mayores preocupaciones de los científicos, pues plantea el problema de la responsabilidad potencial del uso que pueda hacerse de una IA.

En parte debido a esta opacidad, también nos preocupa lo que llamamos el sesgo algorítmico. Los algoritmos, basándose en redes neuronales, a menudo heredan los sesgos de los datos con los que han

sido entrenados: filtrar currículos en la contratación de personal de una empresa por edad, por ejemplo, o calcular la solvencia crediticia de las personas en una entidad bancaria. Estos problemas son difíciles de depurar en la fase de desarrollo y con frecuencia dan lugar a titulares de noticias controvertidos cuando el *software* basado en redes neuronales entra en producción.

No sabemos con exactitud por qué un modelo basado en redes neuronales funciona cuando funciona, pero tampoco por qué falla cuando lo hace. En estos casos, se busca lo que se conoce como *explicabilidad* de la IA. Se trata de un área de investigación que trata de definir métodos y técnicas que expliquen los resultados y soluciones ofrecidos por una IA, de manera que los humanos los puedan entender. Conseguir la explicabilidad de la IA es esencial para poder entender por qué toma una decisión y, de alguna manera, comprender cómo *razona* internamente.

Para agravar más la situación de las llamadas «cajas negras», cuando OpenAI presentó el GPT-4 no dio ningún detalle sobre el tamaño de su red neuronal, el conjunto de datos de entrenamiento, el método de entrenamiento ni de los recursos de computación requeridos. Sin duda, ha sido el anuncio más cerrado de los que ha hecho nunca anteriormente OpenAI (a pesar de que «*open*» quiere decir «abierto»). Esta opacidad sobre el funcionamiento interno de sus IA está empezando a ser habitual entre los gigantes tecnológicos desarrolladores de grandes modelos de lenguaje. Todo indica que la era en que reinaba la cultura de publicar en abierto los hallazgos científicos está llegando a su fin. Lamentablemente, este aislamiento de la información entorpece el desarrollo de las IA y, sobre todo, dificulta la comprensión de aspectos clave (como, por ejemplo, el proceso por el cual una IA toma determinadas decisiones).

Por suerte, hay colectivos de investigadores y desarrolladores que colaboran para conseguir grandes modelos de código abierto más allá de las grandes multinacionales. Stable Diffusion, por ejemplo, genera

imágenes a partir de texto y rivaliza con DALL-E 2 de OpenAI. Otro caso interesante es el de BLOOM, el modelo de lenguaje de código abierto y multilingüe (con el mismo número de parámetros de GPT-3), fue creado por más de un millar de investigadores de IA organizados a través de Hugging Face, la principal plataforma y comunidad de código abierto alrededor de la IA. BLOOM fue entrenado en el supercomputador francés Jean Zay durante casi cuatro meses. La misma plataforma Hugging Face también está trabajando en su propia alternativa abierta a ChatGPT. Iniciativas como estas, además de hacer más explicables las IA, permiten que la agenda de investigación esté un poco menos definida por las grandes empresas.

Uso ético de la IA

La IA ha de utilizarse de manera responsable y beneficiosa para la sociedad en general. En su uso se ha de tener en consideración cuestiones como la privacidad, la seguridad, la discriminación, la transparencia o la responsabilidad social, entre otras.

Para comprender a qué nos referimos con uso ético, recuperemos el ejemplo de los coches sin conductor. ¿Cómo debería actuar un vehículo sin conductor ante un accidente de tráfico con consecuencias mortales? Es decir, ¿cómo debería ser la «moral» que guíe a estos vehículos?

En 2016, los medios de comunicación se hicieron eco del experimento «máquina moral», llevado a cabo por el Instituto de Tecnología de Massachusetts (MIT) para establecer un código ético para los coches. Era una plataforma *online* abierta a todos los usuarios que buscaba recoger la perspectiva humana sobre diversos dilemas morales a los que podían enfrentarse los vehículos sin conductor. Por ejemplo: si en el camino del vehículo se interponen una persona y un animal, ¿a cuál debe salvar el coche en caso de no poder frenar a tiempo? Según los datos aportados, lo moral sería atropellar al animal. Como norma, prima salvar al mayor

número de personas, así que, dado el caso entre escoger la vida del conductor o la de dos peatones, es preferible que el conductor sea quien pierda su vida, puesto que es una sola vida frente a dos. O, pongamos por caso, entre atropellar a un niño o a un anciano, se dirimió que lo moral era salvar al niño.

Otros temas, como la privacidad, son asuntos más delicados que generan más atención en el público general. Actualmente, los algoritmos en que se basan los motores de búsqueda en Internet, o simplemente nuestro teléfono móvil, conocen muy bien, gracias al acceso a cantidades masivas de información que generamos, nuestras preferencias, y pueden llegar a inferir cómo pensamos. Aunque esto sucede en todo el mundo, en algunos países la situación se agrava debido a sus omnipresentes sistemas de vigilancia con reconocimiento facial, entre otras tecnologías basadas en la IA. Esto aumenta enormemente el poder y el alcance de los gobiernos autoritarios y erosiona cualquier expectativa de privacidad personal y, en definitiva, la libertad.

Estos casos se encuadran en lo que llamamos un uso poco ético y confiable de la IA, como demandan las «Directrices éticas para una inteligencia artificial fiable» de la Unión Europea y la «Estrategia Nacional de Inteligencia Artificial» española. Como dice la profesora emérita de la ETH de Zúrich, Helga Nowotny, «están sonando fuertes y omnipresentes alarmas éticas y avisos para que el diseño de la IA sea responsable, beneficioso y alineado con los valores humanos».

La ciberseguridad es desde hace tiempo otra de las áreas de preocupación en torno al desarrollo de las IA, pues supone una de las amenazas más inmediatas a nuestra seguridad general: ciberataques mejorados con la IA en infraestructuras físicas o sistemas críticos, que están cada vez más interconectados y gestionados por algoritmos.

Asimismo, las IA generativas facilitan enormemente la posibilidad de producir noticias falsas que son indistinguibles de la realidad. Ya hemos comentado en este libro las múltiples IA que permiten crear textos,

fotografías, audios y vídeos falsos con apariencia de autenticidad. A finales de marzo de 2023 se publicaron unas imágenes creadas por Eliot Higgins, fundador de la página de periodismo de Bellingcat, de un supuesto arresto de Donald Trump. Las imágenes eran falsas, se habían generado con una IA similar a DALL-E llamada Midjourney, pero su gran calidad las volvía extremadamente realistas (más allá de que tuvieran pequeños fallos).

Las primeras imágenes realistas de personas no reales creadas por una IA datan de 2015, y fueron creadas con redes adversariales generativas (*generative adversarial networks*). Desde entonces, la IA ha avanzado tanto que ahora cualquier persona con acceso a Internet puede generar deliberadamente información falsa con apariencia verídica (y usarla para sus propios intereses). Un ejemplo de esto fue el vídeo del presidente de Ucrania, Volodimir Zelenski, ordenando la rendición de su país justo dos semanas después del inicio del conflicto con Rusia. Solo pasaron unas pocas horas hasta que se esclareció que se trataba de un vídeo falso, pero durante ese tiempo fue sin duda una herramienta muy poderosa para confundir a la población. La IA se utilizó como arma (de desinformación, en este caso) en un conflicto militar, campo en el que, quizá, la amenaza más terrorífica a corto plazo es el uso cada vez más frecuente de vehículos no tripulados.

El desarrollo de sistemas de armas totalmente autónomas, con capacidad de actuar sin necesidad de un humano que dé autorización específica, ha abierto —como no podía ser de otra manera— un fuerte debate por sus implicaciones éticas y legales. Estas armas podrían ser utilizadas en masa para atacar a poblaciones enteras que tendrían dificultades extremas para defenderse. Incluso confiar en que estas armas serán capaces de reconocer su blanco, elegido previamente por un humano, genera inquietudes éticamente perturbadoras. Una parte importante de la comunidad científica lucha apasionadamente por prevenir este uso, e incluso existe una iniciativa en las Naciones Unidas para prohibir este tipo de armamento.

Cuando hablamos de la ética de la IA nos referimos también al conjunto de principios éticos y morales que deben guiar su desarrollo, incluso cuando hablamos de prácticas que corresponden, en realidad, a la ética de las empresas desarrolladoras (más allá de que estén trabajando en IA o en cualquier otro proyecto). Por ejemplo, la IA exige una fuerza laboral humana considerable, puesto que, como hemos visto en el capítulo 5, en el proceso de entrenamiento de la IA generativa la mano humana está muy presente. La revista *Time* informó a principios del año 2023[7] de que muchos de los trabajadores en IA son subcontratados por las grandes empresas tecnológicas en países pobres, lo que se traduce en bajos salarios, condiciones laborales precarias y tareas repetitivas como el etiquetado de datos ante una pantalla.

Impacto social

Al igual que la imprenta, el tren, los transistores y cualquier otra tecnología disruptiva, la IA tiene, y tendrá, un impacto en nuestro día a día. Ya hace tiempo que estamos rodeados de una IA que no vemos, desde la que usa Netflix para recomendarnos qué ver, a la de las empresas energéticas para predecir nuestra demanda de electricidad. Sin embargo, lo más revolucionario aún está por venir, y condicionará nuestras vidas en aspectos mucho más decisivos.

A medida que la IA avanza lo hacen también las oportunidades de usarla para aumentar la productividad y la calidad en muchos más sectores de la economía, incluidos el cuidado de la salud, la educación y el transporte. Aunque es una buena noticia, plantea una amenaza directa: el desempleo tecnológico. En este punto muchos teorizan, pero pocos encuentran realmente cómo afrontar lo que representa uno de los grandes problemas desde el punto de vista socioeconómico.

Es difícil imaginar los cambios que nos esperan en los próximos diez años, y más difícil todavía los que vendrán después. El desarrollo de la

IA conlleva también una enorme (potencial, de momento) capacidad para darle la vuelta tanto al mercado laboral como a la economía global hasta un grado que probablemente no tenga precedentes (ni siquiera la Revolución Industrial). Es una certeza que la IA obligará a reconvertirse a millones de empleados, porque sus funciones serán redundantes e innecesarias.

Diversos estudios apuntan a que la automatización de trabajos impulsada por la IA es una preocupación cada vez más apremiante a medida que la tecnología es adoptada en los distintos sectores económicos. Por ejemplo, un reciente estudio encargado por OpenAI indicó que «aproximadamente el 80 % de la fuerza laboral de Estados Unidos podría ver al menos el 10 % de sus tareas laborales afectadas por la introducción de GPT, mientras que alrededor del 19 % de los trabajadores podrían ver afectadas al menos el 50 % de sus tareas».[8]

Hasta hace pocos años no creíamos que la IA fuese a afectar a los empleos creativos de forma inmediata. Parecía que «solo» estaban en jaque aquellos trabajos rutinarios y repetitivos (los automatizables), como las cadenas de producción o almacenes. Sin embargo, estamos viendo que las tareas creativas también pueden automatizarse con IA para los llamados trabajadores de cuello blanco, es decir, aquellos cuyas tareas implican escribir, asesorar, resumir o resolver problemas complejos. Como apuntó el Premio Nobel de Economía Paul Krugman, la aparición de ChatGPT es solo un ejemplo de esta tecnología que parece capaz de llevar a cabo tareas que hasta hace poco requerían los servicios, no solo de seres humanos, sino también de humanos con una considerable educación académica.

El alcance de la IA sobre los trabajos tradicionalmente más «creativos» es tal que incluso se están sacudiendo los cimientos de la industria audiovisual y editorial. Ya se celebran festivales que exhiben títulos generados con IA, como el Runway AI Film en Nueva York, que tuvo su primera edición en 2023. Y si bien los primeros GPT generaban textos breves e inconexos entre uno y el siguiente, la aparición del GPT-4 ha

constatado que dentro de poco se podrán escribir muchas páginas de forma coherente, estructurada y con un objetivo. Es decir, un libro. Este mismo, por ejemplo.

Es cuestión de tiempo que el discurso de ese político que se ha hecho viral, o ese *bestseller* que vemos en el escaparate de nuestra librería favorita, nos haga preguntarnos si lo ha escrito una persona o una IA, o en cualquier caso, en qué proporción ha participado cada uno. Quizás habrá dos circuitos de libros: los escritos por una IA y los escritos por humanos. El escritor Jorge Carrión, autor de *Los campos electromagnéticos* (un libro que ha escrito usando GPT-2 y 3 en colaboración con el colectivo artístico Estampa), propone en este sentido una analogía interesante: cuando a finales de la Edad Media se pasó de los libros manuscritos a los impresos, la imprenta no suprimió la circulación de libros copiados a mano, que eran más baratos y permitían hacer añadidos o esquivar la censura, sino que se mantuvieron ambos.

La integración de ChatGPT o GPT-4 en los buscadores afectará directamente a muchos negocios digitales cuyos ingresos se generan por el tráfico que dirigen a sus portales y permite la visualización de publicidad. Si GPT-4 nos selecciona la información y nos la ofrece de forma estructurada en un discurso coherente, es posible que afecte a la forma en que navegamos en Internet, pues ya no tendremos necesidad de visitar diversas páginas a través de enlaces de búsqueda hasta llegar a lo que nos interesa.

¿Y se crearán puestos de trabajo nuevos y no automatizados suficientes para absorber a los trabajadores que pierden su trabajo habitual? Y si es así, ¿tendrán estos trabajadores las habilidades, capacidades y rasgos de personalidad necesarios para desempeñar con éxito estos roles de nueva creación? En los últimos años (por no decir meses) ha aparecido un nuevo puesto de trabajo: «ingeniero de instrucciones». Consiste en saber qué instrucciones se han de dar a una IA para lograr la respuesta deseada. Entidades como la World Economic Forum estima que dos de cada tres niños de entre 6 y 12 años tendrán

en el futuro trabajos muy diferentes a los nuestros porque, entre otras cosas, desaparecerán muchísimos de los perfiles laborales que hoy conocemos.

Mi colega Esteve Almirall, profesor de Esade, pronostica que donde más impacto tendrá la IA laboralmente será en el área legal: «Durante décadas el trabajo en los bufetes no ha cambiado sustancialmente. Se basa en buscar precedentes, jurisprudencia, casi todas son tareas similares aunque no idénticas, y requieren capacidad de sintetizar y combinar textos. Precisamente lo que una IA generativa hace tan bien», explica.

Actualmente, el debate se plantea también en el terreno educativo, pues la irrupción de estos actores tecnológicos obligará a rediseñar muchos procesos dentro y fuera del aula. Los ensayos de los estudiantes han sido un instrumento de evaluación desde hace decenios. Ahora, en cambio, cualquier alumno dispone de ChatGPT o de una tecnología equivalente capaz de realizar un trabajo en minutos. Los profesores ya no pueden garantizar que el estudiante sea el autor del ensayo. Se puede decir que se ha democratizado la capacidad de copiar. ¿Deberán los docentes convertirse en inspectores de originalidad abocados a averiguar si ha sido la inteligencia humana o la artificial la que ha resuelto las consignas?

Si bien es cierto que no es la primera vez que el sistema educativo tiene que adaptarse a cambios importantes, no deja de ser un reto. Cuando las calculadoras aparecieron, cambiaron lo que se evaluaba en las clases de matemáticas. Lo mismo ocurrió con la aparición del buscador de Google y Wikipedia, que provocó que la necesidad de memorizar fechas, acontecimientos y datos en general fuera mucho menos importante. Wikipedia externalizó la memoria; ahora, la IA externalizará el procesamiento de la información.

Lo que resulta paradójico es que los mismos ingenieros e ingenieras que programan la IA se verán afectados por ella, ya que parte del código será escrito por una IA y se necesitará mucha menos mano de obra. Los

que sean realmente buenos o tengan más experiencia se reconvertirán en fiscalizadores o supervisores de calidad de lo que produce una herramienta de IA. Conozco ya varios programadores que están usando ChatGPT tanto para que les ayude a crear código como para que lo traduzcan a lenguaje natural.

Me gustaría acabar el capítulo con una visión en positivo de alguien que siempre me ha inspirado: el Dr. Andreu Mas-Colell. Hace unos días leía en su columna semanal de *La Vanguardia* que «las dificultades que genere el Chat podremos superarlas. Y, en cambio, lo que nos puede quedar es un gran empuje hacia la exigencia. ¿Quién quiere gastar tiempo y energía mental en hacer lo que puede hacer una máquina? De cuantas más tareas nos liberen, mas podremos dedicarnos a trabajar en la frontera de lo que ellas pueden hacer. No nos faltará trabajo. La frontera es móvil y no es unidimensional. Es más bien como un globo: cuanto mayor es el globo, mayor es la frontera».

En resumen, ¿nos debe preocupar el impacto de la IA?

- Las IA actuales se consideran «cajas negras», en el sentido de que no tenemos capacidad explicativa para saber cómo llegan a tomar una decisión determinada. Es una de las grandes preocupaciones que despierta la IA.
- El sesgo de una IA no se diseña de forma deliberada, sino que se introduce con datos que ya llevan una carga sesgada.
- Ante el uso de la IA debemos considerar cuestiones como la privacidad, la seguridad, la discriminación, la transparencia o la responsabilidad social, entre otras.
- En el mundo digital la IA ha disparado las opciones para el engaño a través de las noticias falsas, porque la verdad y la falsedad nos llegan por el mismo canal.
- El desarrollo con IA de sistemas de armas totalmente autónomas ha abierto un intenso debate por sus implicaciones éticas y legales.

- Se prevé un cambio de perfiles de puestos de trabajo a causa de la IA que afectará a todos los sectores (incluidas las tareas creativas), ya que la IA es cada vez más capaz de realizar tareas cognitivas complejas.
- La IA tiene el potencial de darle la vuelta tanto al mercado laboral como a la economía global hasta un grado que probablemente no tenga precedentes.

8. ¿Podemos prescindir de la IA?

Aunque no sea unánime, la preocupación de la comunidad científica ante el impacto de la IA tiene fundamento, no solo por el uso malicioso que se le pueda dar sino también porque, además, los modelos actuales son propensos a errores, lo que es una limitación y un problema de gran calado. Y como hemos visto, su simple uso ya conlleva unos cambios sociales sin precedentes a escala mundial, probablemente en todas las dimensiones de la vida humana. Ante este panorama tan incierto, cabe preguntarse por qué deberíamos seguir investigando, desarrollando y aplicando la IA como estamos haciendo hasta ahora. ¿Por qué no detenemos el avance de la IA?

La oportunidad de continuar impulsando la IA

Un breve repaso a los avances de la civilización nos muestra que, en líneas generales, buena parte del progreso humano se debe a que hemos concentrado nuestras inteligencias en tareas que exigían un mayor talento, y hemos mecanizado todo cuanto nos ha sido posible. Ahora, la IA representa una gran oportunidad para ampliar nuestra inteligencia humana, y expandir nuestras capacidades de razonamiento y creatividad para la resolución de problemas complejos.

En este escenario tenemos que aceptar que la IA ha llegado para quedarse. Es una consecuencia más de la automatización —sistémica y

sistemática— en la que se encuentra sumida la humanidad desde hace siglos. Nuestra civilización es el resultado de aplicar la inteligencia humana; ahora podemos tener acceso a una inteligencia mucho mayor, una amplificación de nuestro intelecto y de nuestra creatividad. Cuanto más sagaz sea la IA, más nos permitirá utilizar nuestro cerebro en nuevos campos, nuevas tareas, persiguiendo nuevos límites. Se redefinirá el concepto de inteligencia.

Disponemos de un nuevo medio para abordar los grandes retos a los que nos enfrentamos los humanos. La ciencia, y en particular la inteligencia artificial, puede dar respuesta a los problemas de este siglo, permitiendo fusionar la creatividad humana con la capacidad innovadora de la IA. Debemos aceptar con humildad esta coevolución en la que nos encontramos, pues marca una nueva era de la evolución humana. Si el mundo del futuro tiene alguna posibilidad de ser mejor, lo será también gracias a la participación de la IA, pues su uso aumentará las capacidades de la inteligencia humana. Nuestra principal tarea en esta etapa es establecer entre todos los mecanismos que nos garanticen tener bajo control cualquier potencial uso negativo de las IA.

De hecho, son muchos los trabajos de investigación en la actualidad que buscan favorecer que los sistemas de IA puedan y deban trabajar junto a los humanos para aumentar su inteligencia en lugar de reemplazarlos. Esto permitirá desarrollar equipos humanos-IA de apoyo a las decisiones que puedan, por ejemplo, ayudar tanto a investigadores en medicina personalizada a crear mejores medicamentos como a los propios facultativos a tomar decisiones más acertadas en la prescripción de dichos medicamentos. Nuestro objetivo debe ser formar un equipo que sea capaz de resolver problemas mejor de lo que lo harían los humanos o la IA en solitario.

La IA requiere de una regulación entre todos

Si no podemos prescindir de la IA, ¿cómo podemos prevenir los peligros que implica su uso y evitar los riesgos futuros —que aún no conocemos— y los actuales, los que ya sabemos que están ahí, como permitir que una IA propensa a errores tome cada vez más decisiones? O peor aún, ¿cómo haremos para que esta decisión no quede en manos de unas pocas compañías privadas que deben más lealtad a sus accionistas que a la sociedad?

La mayoría de los expertos coinciden en que, al igual que ha ocurrido en otras áreas importantes en nuestra vida, es necesario establecer un marco regulador consensuado. A medida que avanza el desarrollo de las IA se hace cada vez más evidente la necesidad de que se definan formas de regulación que velen por su desarrollo y uso, en el más amplio sentido del término: leyes, reglas, protocolos de actuación, autorregulación, o sensibilización ciudadana, entre otras.

A nadie le gusta ser regulado, pero todo lo que puede entrañar un potencial peligro para la sociedad lo está. Las aplicaciones específicas de la IA deben ser también reguladas y, en algunos casos, prohibidas si fuese necesario.

Hoy no podemos imaginarnos una sociedad avanzada sin una agencia como la Agencia Europea de Medicamentos, que regula y controla la producción, la distribución y el uso de los fármacos. Lo mismo ha de plantearse en el campo de la IA para garantizar que se desarrolle de forma correcta.

Ahora bien, regular no es sencillo, y se complica mucho más en un mundo globalizado sin un orden político general. Son frecuentes las referencias del secretario general de las Naciones Unidas sobre los beneficios, pero también sobre los riesgos, que la IA puede representar para el mundo, y por ello apela a intentar conseguir un entendimiento mundial sobre el tema, puesto que ningún país por sí solo pueda hacer gran cosa ante este reto. Es crucial que los investigadores y la propia industria del sector colaboren para abordar las implicaciones éticas, sociales y económicas de la IA.

En el ámbito europeo, la Comisión Europea se ha mostrado desde hace tiempo muy activa en materia de regulación, ponderando la observada por Estados Unidos (de menor control) y las de China (de mayor control). En abril de 2021, la Comisión Europea publicó una propuesta de marco jurídico para Europa en la que justificaba la regulación del uso de la IA en sus múltiples beneficios potenciales para la sociedad. Asimismo, indica que dicha regulación no debe olvidar los peligros que entrañan algunos de esos sistemas para evitar situaciones no deseadas. La regulación que se propone para Europa persigue facilitar la innovación de la IA definiendo varios niveles de protección como la seguridad y la defensa de los derechos fundamentales y la creación de un ecosistema de confianza. Presenta, además, requisitos específicos tanto para el desarrollador de la IA como para el usuario durante el ciclo de vida de la IA. La atención se centra en los aspectos éticos, legales y técnicos de su uso: la idea es responsabilizar a las empresas de las infracciones de privacidad o de las decisiones injustas tomadas por sus IA. Para las aplicaciones de IA de alto riesgo existen requisitos adicionales.

A finales de 2022, los Estados miembros de la Unión Europa aprobaron una primera versión del Reglamento de Inteligencia Artificial de la Comisión Europea (el *Artificial Intelligence Act*) y a partir de aquí se espera que entre finales de 2023 y principios de 2024 el Parlamento Europeo, la Comisión Europea y los Estados miembros establezcan las negociaciones necesarias para que una Ley de Inteligencia Artificial a nivel europeo pueda ser aplicable en todo el continente lo antes posible.

En esta línea, España ya cuenta con una Secretaría de Estado de Digitalización e Inteligencia Artificial ubicada jerárquicamente dentro del Ministerio de Asuntos Económicos y Transformación Digital, y que dispone de una «Estrategia Nacional de Inteligencia Artificial». En diciembre de 2022 se creó la Agencia Estatal para la Supervisión de la Inteligencia Artificial, cuya función es supervisar y orientar a las empresas en el desarrollo y el uso de la IA, así como penalizar

infracciones de las leyes y reglamentos regulatorios de la materia que deberían ir desarrollándose y poniendo en funcionamiento durante los próximos años.

Pero ¿hasta qué punto será suficiente esta regulación si no es a escala mundial? Recientemente, las capacidades de las IA generativas han avanzado de una manera tan vertiginosa que nos ha cogido a todos por sorpresa. Y como era de esperar, la preocupación de los expertos ha crecido a la misma velocidad. Sin ir más lejos, Geoffrey Hinton —el más reconocido investigador actual en el ámbito de la IA y quien fuera director de aquel equipo de la Universidad de Toronto que cambió el curso de la Historia en el concurso ImageNet de 2012— decidió en mayo de 2023 abandonar Google para poder hablar sin ataduras de los peligros de la IA. Una de sus preocupaciones es que el uso masivo de estas nuevas IA generativas provoque un aluvión de material audiovisual falso tan sofisticado que la población no pueda distinguir entre lo que es verdadero y lo que no lo es. Es decir, la verdad podría llegar a convertirse en un concepto fácilmente manipulable y pervertido.

Pero, además, Geoffrey Hinton expone su preocupación por las consecuencias de la competencia agresiva entre las grandes tecnológicas, esta carrera por la IA que se aceleró cuando OpenAI lanzó ChatGPT en noviembre de 2022. Los tiempos son significativos: solo dos meses más tarde, Microsoft anunció la integración de esta IA en su buscador de Internet Bing, mientras Google anunció que usaría una alternativa llamada Bard en su navegador en respuesta al desafío. Para Geoffrey Hinton, imaginar el camino que esto puede seguir «da miedo».

Somos muchos los expertos que compartimos la preocupación de Hinton. La mayoría estamos de acuerdo con los grandes nombres que firmaron la carta «Pause Giant AI Experiments» mencionada en el capítulo 6: es necesario avanzar con cautela y ralentizar los desarrollos de la IA basada en la fuerza bruta hasta que se logre una mejor comprensión de su potencial impacto negativo. Y aprovechar esa desaceleración para, al mismo tiempo, avanzar en la regulación. Este es

un reto difícil porque requiere la colaboración de científicos, gobiernos, corporaciones tecnológicas y la sociedad en general. Y dado lo rápido que se mueven las cosas, no hay tiempo que perder: necesitamos una respuesta inmediata a nivel mundial, precisamente porque la carrera emprendida entre las grandes tecnológicas es a nivel global. No lo olvidemos.

Soberanía europea

A estas alturas, el lector ya tendrá claro que la IA en general, y las IA generativas en particular, serán una fuente de enorme poder para las empresas y los países y que pronto asistiremos a una carrera para su adopción, pese a que nadie tiene idea del verdadero impacto que puede representar. Sin embargo, ello no es excusa para no hacernos preguntas: si estas IA generativas son tan importantes, ¿cómo es que ahora mismo estamos cediendo su control a unos laboratorios privados estadounidenses?

Las IA generativas como las que están a disposición de cualquier usuario hoy en día tienen que considerarse no solo como un artefacto técnico, sino también como un artefacto político. Especialmente si siguen controladas por unos pocos actores del sector privado (GPT de OpenAI con el respaldo de Microsoft, LaMDA de Google o LLaMA de Meta). Actualmente, las IA generativas están controladas mayoritariamente por laboratorios y empresas no europeas. Los modelos de lenguaje de código abierto europeos, como BLOOM o el Aleph-Alpha alemán, están todavía lejos de compararse con los modelos de las grandes tecnológicas americanas ya disponibles.

Como ya hemos dicho, se necesita una supercomputación a gran escala para su creación, algo que resulta muy costoso y que solo está al alcance de unos pocos. Por el momento, todo indica que el número de estas IA generativas en los próximos años continuará siendo

relativamente limitado.

¿Tiene sentido que Europa no reaccione y espere a que estas IA y sus propietarios se vuelvan más poderosos y dominantes? ¿Por qué no decide crear sus propias IA generativas y así no permanecer a merced de estas contadas empresas no europeas?

Europa necesita tener control sobre estas tecnologías que serán tan nucleares en el futuro de nuestra sociedad y economía. En realidad, es una situación de dependencia que existe desde hace muchos años, pero que ahora se está acelerando. El Dr. Josep Lluís Berral, miembro del BSC, estima que «la apuesta se ha doblado», en referencia a que en Europa llevamos dos decenios dependiendo de un algoritmo opaco y controlado por una empresa californiana (Google) cuando hacemos algo tan básico pero esencial como buscar algo en Internet, mientras que otros actores ya han desarrollado sus propios buscadores: Rusia con Yandex y China con Baidu.

Pero ¿tenemos la capacidad computacional necesaria para hacerlo posible? En estos momentos, contamos en Europa con tres supercomputadores a gran escala: el MareNostrum 5 en Barcelona, uno en Kajaani (Finlandia) y otro en Bolonia (Italia). Todos ellos tienen actualmente (por parte de científicos y de empresas) una demanda de uso tal que hay listas de espera. De modo que si pretendemos tener la soberanía europea en cuanto a IA, necesariamente debemos incrementar la capacidad de supercomputación a gran escala a nivel europeo.

Ya he citado antes a Mateo Valero, director de BSC, y vuelvo a recurrir a él y a una reflexión que lleva años pregonando: Europa debe diseñar y fabricar chips europeos. Entre otras cosas, para crear estos supercomputadores a gran escala con tecnología propia y no depender, nuevamente, de empresas californianas como Intel o Nvidia. Esperemos que los esfuerzos que ya se han iniciado en Europa en este sentido consigan sus objetivos lo más pronto posible.

Este aumento de la capacidad de supercomputación en Europa (que ahora mismo está en un nivel muy inferior de lo que requiere un mundo

digital como en el que ya estamos inmersos) deberá complementarse, en un siguiente paso, con un servicio propio de IA, que se convertirá pronto en una *utility* (servicio público) más y que podrá suministrarse de una manera comparable a la energía eléctrica, el gas o el agua corriente: simplemente conectamos los aparatos a la red eléctrica para hacerlos funcionar. No sabemos qué hay al otro lado del enchufe, ni si la energía es generada por medios eólicos o nucleares. Solo sabemos que podemos hacer uso de un servicio que está disponible cuando lo necesitamos.

En resumen, ¿podemos prescindir de la IA?

- Nuestra civilización es el resultado de aplicar la inteligencia humana. La IA representa una gran oportunidad para ampliar nuestra inteligencia y expandir nuestras capacidades de razonamiento y creativas para la resolución de problemas complejos.
- Para prevenir riesgos urge establecer un marco de regulación que vele por el desarrollo y el uso de la IA, en el más amplio sentido del término. Europa está en proceso de aprobación del reglamento de la IA.
- Para Europa es una prioridad ser soberana en cuanto a la IA, lo que implica, entre otras cosas, incrementar la capacidad de supercomputación a gran escala a nivel europeo.
- La IA se convertirá en un servicio público más, como lo es el acceso al agua, el suministro de energía o el mismo *cloud computing*.

Palabras finales

Creo que no exagero al decir que actualmente la población en general tiene la percepción de que la IA ha alcanzado competencias casi humanas. La realidad es que los algoritmos en que se basan las IA no tienen inteligencia, sino solo «habilidades sin comprensión», en el sentido de que son algoritmos que pueden llegar a ser muy hábiles realizando tareas específicas pero sin comprender absolutamente nada de lo que están haciendo, ni cuentan con el llamado sentido común de la inteligencia humana, que la hace tan diversa, rica, espontánea y maravillosamente imprevisible.

Para que las IA alcancen un nivel general similar al humano todavía hace falta desarrollar nuevas formas de aprendizaje que no requieran de enormes cantidades de datos para ser entrenadas. Dicho de otro modo, para parecerse a la inteligencia humana las IA deberían al menos acercarse a la manera en que aprendemos los humanos, que es mucho más eficiente.

Hemos de asumir, también, que el crecimiento exponencial de la capacidad de computación que hemos visto en los últimos decenios ha llegado a su fin. Los supercomputadores a gran escala ya no podrán evolucionar y seguir creciendo, al menos no de la manera en que lo han hecho hasta ahora (lo que llamábamos crecimiento por fuerza bruta). Principalmente, deberán ser mucho más eficientes en consumo energético, una de las barreras principales al avance de la IA. En comparación, el cerebro humano es mucho más eficiente que los supercomputadores a gran escala y los algoritmos que dan vida a la IA

actual.

En cualquier caso, a pesar de las limitaciones de la actual IA, no debemos olvidar que estamos ante un nuevo paradigma coevolutivo en el que nos hemos embarcado la humanidad junto a la IA. Y esta interdependencia seguirá creciendo, sin duda alguna, porque la búsqueda de una IA General es el Santo Grial de la ciencia.

La IA (incluso en su incipiente estado actual) nos abre a nuevos horizontes, pero también importantes riesgos y problemas. Por ello, hemos de ser muy conscientes de su inevitabilidad y rápida penetración y de que urge establecer una regulación adecuada para evitar las consecuencias negativas (o imprevistas) de una tecnología tan disruptiva. Esta herramienta regulatoria de la que nos dotemos tiene que estar preparada para ir mutando al compás de los avances de la IA y su transición progresiva y silenciosa. Y así, si llega el día en que este viaje nos lleva a algo parecido a una IA General, no nos pillará desprevenidos.

Sin embargo, hemos de tener claro que la mayor o menor rigidez de un reglamento no será suficiente para prevenir los problemas. Mi experiencia me dice que es imprescindible educar a la ciudadanía para dotarla de las competencias necesarias para controlar a la IA (y a su entorno de empresas e instituciones), en vez de ser controlados por ellas. Necesitamos estar todos más informados acerca de su evolución para tener más capacidad crítica y participativa y así contribuir activamente a poner los límites oportunos.

Sinceramente, pienso que no debemos crearnos falsas expectativas con respecto a la IA, al menos a corto plazo. Las sociedades humanas son estructuras muy complejas y no hay soluciones fáciles a los problemas que vendrán. No hay atajos (y tampoco trampas en el camino). El futuro es una elección que tomamos conjuntamente y no algo que simplemente sucede. Todos los países y actores involucrados deberían tener la posibilidad de aportar su visión, y en modo alguno deberíamos dejar las decisiones en manos de un reducido grupo de corporaciones e instituciones.

Espero que el lector haya encontrado en este libro un compañero que lo acerque un poco más a la IA. Es el primer paso para participar en las decisiones colectivas que debemos tomar ante la inquietante, prometedora y casi misteriosa revolución que supone la IA. Para mí, la mayor transformación de la historia de la humanidad.

Agradecimientos

Este libro no habría visto la luz sin el apoyo incondicional de mi familia. En especial, quiero agradecer a Roser y a Júlia su ánimo y apoyo para no desfallecer durante todo el proyecto, y a Bernat por inspirarme a explorar el emocionante mundo de las IA generativas en sus inicios.

La investigación es un deporte de equipo, y sin los brillantes colegas con quien me he cruzado a lo largo de mi carrera académica nunca habrían llegado las aportaciones de este libro. Investigadores como el Dr. Ricard Gavaldà, el Dr. Jordi Nin o el Dr. Joan Capdevila despertaron en mí el interés por los temas de aprendizaje automático. También, los colegas con quienes he desarrollado investigación conjunta en IA que me ayudaron a comprender y reflexionar, como el Dr. Xavier Giró-i-Nieto, la Dra. Miriam Bellver, el Dr. Victor Campos, la Dra. Amanda Duarte, Juan Luis Domínguez o Laia Tarrés.

Son varios los que me hicieron comprender que tenía entre manos un mensaje de interés y valía la pena escribir un libro sobre este tema para compartirlo con el público general. El primero fue Màrius Mollà, pero también Marc Melillas y Pep Martorell.

Mi agradecimiento más sincero a aquellas personas que me proporcionaron valiosas discusiones y sugerencias para mejorar el contenido de este libro: Rubèn Tous, Ramon Canal, Alberto Gutiérrez, Josep Lluís Berral, Albert Viñals, Helena Gamero, Nuria Noriega, Alexandre Puerto y Marta Rosell. No sería justo olvidarme de Jose, Jaime y toda la tropa del bar de la FIB, quienes mantienen ese rincón de la universidad donde la creatividad fluye y ha sido de gran ayuda en

muchas de mis conversaciones con colegas.

Un gran número de personas me han ayudado a encontrar el registro adecuado para dirigirme a un público general en este libro. La complicidad con Xavi Martínez, Rosa de Diego y todos los invitados que han pasado por nuestro programa de IA en Ràdio 4 de RNE ha sido primordial. Sin olvidar a los compañeros y las compañeras con quienes he tenido el honor de compartir el reto de preparar la exposición de IA en el Centre de Cultura Contemporània de Barcelona: Cira Pérez, Carlota Broggi, Lluís Nacenta, Àlex Papalini, Jordi Costa y Judit Carrera.

Especial mención se merece mi editor Ignacio Fernández, insustituible para refinar mi argumentación y que el manuscrito adoptara la forma óptima. Y a Jordi Nadal, por concederme el honor de ser incluido en la lista de autores de la editorial Plataforma.

Finalmente, mi mayor agradecimiento a mi universidad, la Universitat Politècnica de Catalunya, y a mi centro de investigación Barcelona Supercomputing Center. Un entorno de excelencia que me ha permitido realizar mi investigación sobre estos temas y acumular los conocimientos que comparto en este libro. Y, entre todos, quiero destacar a los tres profesores más excelentes que me han apoyado desde mis inicios hace muchos años, el Dr. Mateo Valero, el Dr. Jesús Labarta y el Dr. Eduard Ayguadé. Sin el soporte que siempre me han prestado —y, sobre todo, su paciencia para con mi persona—, nunca habría podido explorar e iniciarme en el el apasionante mundo de la inteligencia artificial.

Notas

1\.

Inteligencia artificial de la mano del Barcelona Supercomupting Center.

2\.

Darrach, B. y Shaky, M. (1970). «The first electronic person», *Life Magazine*, p. 68.

3\.

Silver, D., Schrittwieser, J., Simonyan, K. *et al*. (2017). «Mastering the game of Go without human knowledge», *Nature,* n.º 550, p. 354-359.

4\.

Bostrom, N. (2016). *Superinteligencia. Caminos, peligros, estrategias*, TEELL Editorial.

5\.

VV.AA., (2022). «PaLM: Scaling Language Modeling with Pathways». URL arXiv preprint arXiv:2204.02311

6\.

Russell, S. y Norvig, P. (2004). *Inteligencia artificial*. *Un enfoque moderno*, Pearson.

7\.

Perrigo, B. (2023). «Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic», *Time Magazine*.

8\.

Eloundou, T., Manning, S., Mishkin, P. y Rock, D. (2023). «GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models», Cornell University. URL arXiv preprint arXiv:2303.10130.

Su opinión es importante.
En futuras ediciones, estaremos encantados
de recoger sus comentarios sobre este libro.

Por favor, háganoslos llegar a través de nuestra web:

www.plataformaeditorial.com

Para adquirir nuestros títulos,
consulte con su librero habitual.

«*I cannot live without books*.»
«No puedo vivir sin libros.»
THOMAS JEFFERSON

Desde 2013, Plataforma Editorial planta un árbol
por cada título publicado.

